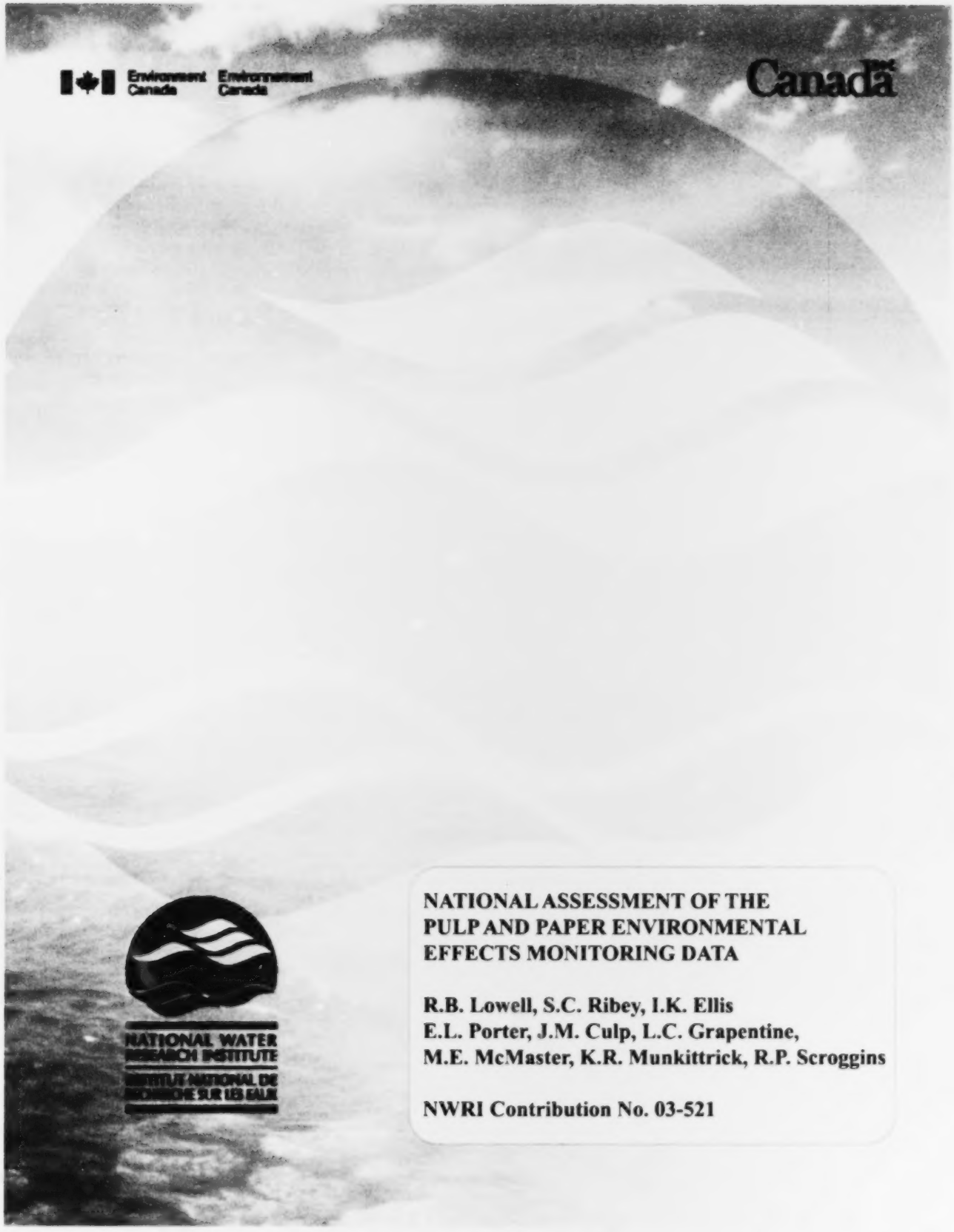


NATIONAL ASSESSMENT OF THE PULP AND PAPER ENVIRONMENTAL EFFECTS MONITORING DATA

R.B. Lowell, S.C. Ribey, I.K. Ellis
E.L. Porter, J.M. Culp, L.C. Grapentine,
M.E. McMaster, K.R. Munkittrick, R.P. Scroggins

NWRI Contribution No. 03-521

103-05636/

National Assessment of the Pulp and Paper Environmental Effects Monitoring Data

R. B. Lowell, S. C. Ribey, I. K. Ellis, E.L. Porter, J. M. Culp, L.C. Grapentine, M.E. McMaster, K. R. Munkittrick, R.P. Scroggins

NWRI Contribution 03-521

April 2003

Abstract

In Canada, all pulp and paper mills need to conduct environmental effects monitoring (EEM) as part of the requirements of the *Pulp and Paper Effluent Regulations* under the *Fisheries Act*. EEM studies typically consist of a benthic invertebrate survey, a fish population survey, and an analysis of possible impacts on fish usability. This assessment of Cycle 2 Environmental Effects Monitoring (EEM) data was conducted to determine the type and magnitude of effects occurring on fish and benthic invertebrate communities on a national scale. The predominant response pattern observed in benthic invertebrate communities across the country was one of mild to moderate eutrophication, particularly in riverine habitats, although more pronounced eutrophication was observed at a few mills. These response patterns were characterized by increases in abundance and either increases or decreases in taxon richness. In contrast, the average response pattern observed in marine/estuarine habitats was more indicative of toxic or smothering effects (decreases in both abundance and taxon richness). The benthic invertebrate response in lakes was intermediate between those in riverine and marine/estuarine habitats, although there were indications of large changes in community composition. The predominant response pattern seen in fish across the country was a decrease in gonad weight and increases in liver weight, condition, and weight at age. This is believed to be indicative of some form of metabolic disruption in combination with a nutrient enrichment effect. With respect to gender, male and female fish showed similar responses for all endpoints, with greater decreases in gonad weight being observed for female fish. The metabolic disruption/nutrient enrichment observed in fish overlapped the eutrophication results seen in benthos. The sublethal toxicity data showed a national improvement in effluent quality at Canadian pulp and paper mills between Cycles 1 and 2. This improvement is largely attributed to the installation of secondary treatment systems after Cycle 1 at many mills across the country. In summary, effluent quality has vastly improved since the PPER were promulgated; however, effects on benthic invertebrates and fish are still being seen.

Résumé

Au Canada, toutes les fabriques de pâtes et papiers sont tenues de faire un suivi des effets sur l'environnement (SEE) dans le cadre des exigences du *Règlement sur les effluents des fabriques de pâtes et papiers* (REFPP) pris en application de la *Loi sur les pêches*. Le SEE consiste généralement en une étude sur les invertébrés benthiques, une étude sur les populations de poissons et une analyse des effets possibles sur la facilité d'utilisation du poisson. Cette évaluation des données du cycle 2 du SEE a été effectuée pour déterminer le type et l'ampleur des effets se produisant sur le poisson et les communautés d'invertébrés benthiques à l'échelle nationale. Le type de réponse prédominant observé chez les communautés d'invertébrés benthiques de tout le pays était une eutrophisation de faible à modérée, notamment dans les habitats fluviaux, mais une eutrophisation plus marquée a été constatée au voisinage de quelques fabriques. Ces types de réponse étaient caractérisés par des augmentations de l'abondance des taxons et des augmentations ou des diminutions de la richesse taxonomique. Par contre, le type de réponse moyen observé dans les habitats marins ou estuariens indiquait plutôt des effets toxiques ou d'étouffement (des diminutions à la fois de l'abondance et de la richesse taxonomiques). La réponse des invertébrés benthiques dans les lacs était intermédiaire entre celles dans les habitats fluviaux et les habitats marins ou estuariens, mais certains signes indiquaient d'importants changements dans la composition des communautés. Le type de réponse prédominant observé chez les poissons partout au pays était une diminution du poids des gonades ainsi qu'une augmentation du poids du foie, de l'adiposité et du poids selon l'âge, ce qui semble indiquer une forme de perturbation métabolique conjuguée à un effet d'enrichissement en substances nutritives. Pour ce qui est du sexe, les réponses des poissons mâles et femelles étaient semblables pour tous les indicateurs, mais on a observé que le poids des gonades diminuait davantage chez les femelles. La perturbation métabolique et l'enrichissement en substances nutritives observés chez le poisson chevauchaient les résultats de l'eutrophisation constatés chez le benthos. Les données des études de toxicité sublétale ont montré que la qualité des effluents des fabriques de pâtes et papiers au Canada s'était améliorée à l'échelle nationale entre les cycles 1 et 2, surtout en raison de l'installation, après le cycle 1, de systèmes de traitement secondaire dans un grand nombre de fabriques au pays. En résumé, la qualité des effluents s'est considérablement améliorée depuis que le REFPP a été promulgué; toutefois, des effets sur les invertébrés benthiques et le poisson sont encore constatés.

Report Authors

Dr. Richard Lowell	National EEM Office, National Water Research Institute (NWRI), Environment Canada
Sandra Ribey	National EEM Office, NWRI, Environment Canada
Ingrid Ellis	National EEM Office, NWRI, Environment Canada
Edward Porter	National EEM Office, NWRI, Environment Canada
Dr. Joseph Culp	National Water Research Institute, Environment Canada
Dr. Lee Grapentine	National Water Research Institute, Environment Canada
Dr. Mark McMaster	National Water Research Institute, Environment Canada
Dr. Kelly Munkittrick	University of New Brunswick
Rick Scroggins	Environmental Technology Centre, Environment Canada

Acknowledgements

The authors extend special thanks to the following people for reviewing and providing constructive comments on various drafts of this document: Dr. Olof Sandström (SKUTAB Coastal Development Ltd., Sweden), Dr. Robert Bailey (University of Western Ontario), Dr. Michael Paine (Paine, Ledge and Associates), Dr. Simon Courtenay (Department of Fisheries and Oceans, Moncton), Nancy Glozier (Ecosystem Health Assessment, Environment Canada, Prairie and Northern Region), Raymond Chabot (Environment Canada, Quebec Region), Dr. Alan Willsie (Environment Canada, Quebec Region), Kathleen Hedley (National EEM Office, NWRI, Environment Canada), Lise Trudel (National EEM Office, NWRI, Environment Canada), and Sherry Walker (National EEM Office, NWRI, Environment Canada). We also thank the EEM National Team and Science Committee for their input throughout this study.

We would also like to thank the following people for their support with data quality assurance/quality control and statistical analysis: Marc Charette (National EEM Office, NWRI, Environment Canada), Jason Inkster (National Water Research Institute, Environment Canada), Gilles Champagne (National EEM Office, NWRI, Environment Canada), Dr. Georgine Pastershank (National EEM Office, NWRI, Environment Canada), and Dr. Bruce Kilgour (Jacques Whitford Environment Limited). The authors also acknowledge Marla Sheffer for her work as editor.

Executive Summary

The *Pulp and Paper Effluent Regulations* (PPER) requires Canada's pulp and paper mills to conduct Environmental Effects Monitoring (EEM) studies on their receiving environments in order to assess and monitor the potential effects of their effluents on fish, fish habitat, and the use of fisheries resources. Since the EEM program's inception in 1992, pulp and paper mills have completed two three- to four-year sequences of monitoring and interpretation phases known as "cycles." The EEM studies focus mainly on fish population surveys and benthic invertebrate community surveys; also, where required supporting information pertaining to the sublethal toxicity of effluents, water and sediment quality, and fish tissue is collected.

The following report provides a national assessment of EEM data and the target audience is the scientific community. This assessment attempts to determine the type and magnitude of effects occurring on fish and benthic invertebrate communities. Environment Canada scientists guided the data analysis through the development of a series of key questions or priority areas of focus. The questions mainly focused on the types and magnitude of effects occurring in fish and benthic invertebrate communities. These effects were evaluated through a combination of ANOVA, ANCOVA, and meta-analyses. The unique combination of statistical approaches, which synthesized the results from large numbers of mills, facilitated the addressing of questions that have received very little study to date. The assessment is fundamental to evaluating the effectiveness of EEM monitoring approaches in assessing the effects of pulp and paper mill effluents on aquatic environments.

The focus of this national assessment is placed on Cycle 2 EEM data, although, where possible, some comparisons are made with Cycle 1 results. In the case of the benthic invertebrate community survey, of the reports that noted changes between Cycles 1 and 2, approximately three-quarters of the mill studies reported an improvement in benthic conditions between cycles. Nevertheless, effects of pulp mill effluent on benthic invertebrate communities were still widely present during Cycle 2, although apparently less pronounced than in previous years. The benthic invertebrate community survey data were analyzed, in part, to determine the influence of habitat type, effluent treatment and mill process type, and study design on invertebrate community response to effluent exposure. The predominant response pattern observed in benthic invertebrate communities across the country was one of mild to moderate eutrophication, particularly in riverine habitats, although more pronounced eutrophication was observed at a few mills. These response patterns were characterized by increases in abundance and either increases or decreases in taxon richness. The eutrophication was likely due to the nutrient enrichment effects of the phosphorus, nitrogen, and organic content of the effluent. In contrast, the average response pattern observed in marine/estuarine habitats was more indicative of toxic or smothering effects (decreases in both abundance and taxon richness). The benthic invertebrate response in lakes was intermediate between those in riverine and marine/estuarine habitats, although there were indications of large changes in community composition. These response patterns were national averages, and exceptions occurred in each of the habitat subgroups. The two most common effluent

treatment types in Canada are activated sludge treatment and aerated stabilization basins. Similar mild eutrophication effects were observed for both treatment types. The two most common mill process types are kraft and thermomechanical, and the results suggested a milder eutrophication response for kraft mills. The analyses further provided for a number of recommendations for improvements in study design for the benthic invertebrate community survey.

Data from the fish population survey were analyzed to determine the magnitude of effects on the key fish endpoints as influenced by gender, fish species, effluent treatment, mill process type, and habitat. The predominant response pattern seen in fish across the country was a decrease in gonad weight and increases in liver weight, condition, and weight at age. This is believed to be indicative of some form of metabolic disruption in combination with a nutrient enrichment effect. Although the mechanism of such responses and their ecological importance remain unclear, these patterns of effects on a national scale may be related to some form of disruption of endocrine functioning, which is key to growth and reproductive development. Although this combined response pattern (metabolic disruption and nutrient enrichment) was the one that was most commonly observed, effluent-exposed fish at other mills experienced increases in all of the three main endpoints — gonad weight, liver weight, and condition factor — which is often an indication of an increase in nutrient availability. This overall nutrient enrichment may sometimes mask any metabolic disruption that may be occurring. A third major response pattern that was observed was indicative of nutrient limitation or chemical toxicity. This latter response was generally associated with decreases in most or all of the endpoints. The general patterns observed in Cycle 2 were consistent with what was observed in Cycle 1. With respect to gender, male and female fish showed similar responses for all endpoints, with greater decreases in gonad weight being observed for female fish. The two most commonly used species nationally were white sucker and longnose sucker, and the response patterns for these two species were consistent with the national average effect (metabolic disruption combined with nutrient enrichment). Similar responses were observed for activated sludge treatment versus aerated stabilization basins, as well as for kraft versus thermomechanical process types, with all being consistent with the national average response pattern. There were some indications of different responses among the major habitat types, however, since most of the fish studies were conducted in riverine habitats, a more detailed evaluation of the influence of habitat type awaits the availability of further data.

The meta-analytical results from the benthic invertebrate and fish population surveys were quantitatively integrated using multidimensional scaling ordination. The overall response patterns showed good agreement between the fish and benthic results. The metabolic disruption/nutrient enrichment observed in fish overlapped the eutrophication results seen in benthos. The results also revealed that overall fish nutrient enrichment (without measurable metabolic disruption) may not be manifested until the benthos reach a state of pronounced eutrophication. Moreover, the fish results indicated the presence of metabolic disruption effects even when the overall effect was one of eutrophication. In summary, the core components of the EEM field survey (fish and benthic invertebrate

community surveys) complemented each other by revealing different aspects of the effects associated with effluent exposure.

The sublethal toxicity data showed a national improvement in effluent quality at Canadian pulp and paper mills between Cycles 1 and 2. This national trend for improvement was particularly clear in toxicity levels for *Ceriodaphnia*, fathead minnow, *Selenastrum*, and inland silverside. Results of tests with echinoderms and *Champia parvula* showed less striking tendencies for lowered toxicity in Cycle 2. Decreases in effluent toxicity between Cycle 1 and Cycle 2 are largely attributed to the installation of secondary treatment systems at mills across the country.

In summary, effluent quality has vastly improved since the PPER were promulgated; however, effects on benthic invertebrates and fish are still being seen. More monitoring is needed to assess the spatial extent and ecological importance of these observed effects, and Environment Canada will continue to work with industry and other stakeholders to better understand their significance.

Sommaire

Le Règlement sur les effluents des fabriques de pâtes et papiers (REFPP) exige des fabriques canadiennes qu'elles réalisent des études de suivi des effets sur l'environnement (SEE) portant sur leurs milieux récepteurs afin d'évaluer et de surveiller tout effet possible de leurs effluents sur le poisson, l'habitat du poisson et l'utilisation des ressources halieutiques. Depuis le début du programme de SEE en 1992, les fabriques de pâtes et papiers ont complété deux, trois ou quatre séries annuelles de phases de suivi et d'interprétation connues sous le nom de « cycles ». Les études de SEE sont surtout axées sur les populations de poissons et les communautés d'invertébrés benthiques, et au besoin, des données justificatives sur la toxicité sublétale des effluents, la qualité de l'eau et des sédiments et les tissus des poissons sont aussi recueillies.

Le rapport suivant présente une évaluation nationale des données du SEE à l'intention de la communauté scientifique. Cette évaluation s'efforce de déterminer le type et l'ampleur des effets qui se produisent chez les poissons et les communautés d'invertébrés benthiques. Les scientifiques d'Environnement Canada ont piloté l'analyse des données par une série de questions clés ou de zones d'attention prioritaires. Les questions mettaient l'accent principalement sur le type et l'ampleur des effets qui se produisent chez les poissons et les communautés d'invertébrés benthiques. Ces effets ont été évalués par une combinaison d'analyses de la variance, d'analyses de la covariance et de méta-analyses. L'unique combinaison d'approches statistiques, qui synthétisait les résultats provenant d'un grand nombre de fabriques, a facilité la formulation des questions n'ayant été que très peu étudiées à ce jour. L'interprétation est essentielle à la détermination de l'efficacité des méthodes de SEE pour l'évaluation des effets des effluents des fabriques de pâtes et papiers sur les milieux aquatiques.

Cette évaluation nationale est axée sur les données du deuxième cycle du SEE, même si, lorsque c'était possible, quelques comparaisons sont faites avec les résultats du premier cycle. Pour ce qui est des communautés d'invertébrés benthiques, les trois quarts environ des rapports ayant signalé des changements entre le premier et le deuxième cycle ont mentionné une amélioration des conditions benthiques entre les cycles. Cependant, les effets des effluents des fabriques de pâtes sur les communautés d'invertébrés benthiques étaient encore largement présents pendant le deuxième cycle, quoique, en apparence, ils étaient moins prononcés que dans les années antérieures. Les données concernant les communautés d'invertébrés benthiques ont été analysées, en partie afin de déterminer l'influence du type d'habitat, du type de traitement de l'effluent de la fabrique, du type de procédé de la fabrique, et du modèle d'étude utilisé pour observer les réponses des communautés d'invertébrés à l'exposition à un effluent. Le type de réponse prédominant observé chez les communautés d'invertébrés benthiques de tout le pays était une eutrophisation de faible à modérée, particulièrement dans les habitats fluviaux, même si une eutrophisation marquée a été constatée au voisinage de quelques fabriques. Ces types de réponses étaient caractérisés par des augmentations de l'abondance des taxons et par des augmentations ou des diminutions de la richesse taxonomique. L'eutrophisation était probablement due aux effets de l'enrichissement en substances nutritives causé par le

phosphore, l'azote et les matières organiques de l'effluent. Par contre, le type moyen de réponse observé dans les habitats marins ou estuariens indiquait plutôt des effets toxiques ou d'étouffement (diminutions à la fois de l'abondance et de la richesse taxonomiques). La réponse des invertébrés benthiques dans les lacs était intermédiaire entre celles observées dans les habitats fluviaux et les habitats marins ou estuariens, mais certains signes indiquaient d'importants changements dans la composition des communautés. Ces types de réponse étaient des moyennes nationales, et des exceptions sont survenues dans chacun des sous-groupes d'habitats. Les deux types de traitement des effluents les plus courants au Canada sont le traitement aux boues activées et l'utilisation d'étangs de stabilisation aérobie. De faibles et similaires effets d'eutrophisation ont été observés pour les deux types de traitement. Les deux types les plus courants de procédé employés par les fabriques sont les procédés kraft et thermomécanique, et les résultats portaient à croire que l'eutrophisation était plus faible dans les fabriques de pâte kraft. Les analyses ont également permis d'établir un certain nombre de recommandations concernant les améliorations à apporter au plan d'étude des communautés d'invertébrés benthiques.

Les données des études des populations de poissons ont été analysées afin de déterminer l'ampleur des effets sur les principaux indicateurs chez le poisson en fonction du sexe et de l'espèce, du type de procédé employé par les fabriques, du traitement des effluents, et de l'habitat. Le type de réponse prédominant observé chez les poissons partout au pays était une diminution du poids des gonades ainsi qu'une augmentation du poids du foie, du facteur de condition et du poids selon l'âge, ce qui semble indiquer une forme de perturbation métabolique conjuguée à un effet d'enrichissement en substances nutritives. Bien que le mécanisme et l'importance écologique de ces réponses demeurent incertains, ces types d'effet, à l'échelle nationale, peuvent être reliés à une certaine forme de perturbation de la fonction endocrinienne, qui est essentielle à la croissance et au développement reproductif. Même si ce type de réponse conjuguée (perturbation métabolique et enrichissement en substances nutritives) était le plus fréquemment observé, les poissons exposés à l'effluent d'autres fabriques ont connu des augmentations en ce qui concerne chacun des trois principaux indicateurs : le poids des gonades, le poids du foie et le facteur de condition, ce qui est souvent une indication d'une augmentation de la disponibilité des substances nutritives. Cet enrichissement global en substances nutritives peut parfois masquer toute perturbation métabolique susceptible de se produire. Un troisième type important de réponse qui a été observé indiquait une limitation en substances nutritives ou une toxicité chimique. Cette dernière réponse était généralement associée à des diminutions de la plupart ou de la totalité des indicateurs. Les tendances générales observées dans le deuxième cycle concordaient avec ce qui avait été observé dans le premier cycle. Pour ce qui est du sexe, les réponses des poissons mâles et femelles étaient semblables pour tous les indicateurs, mais la diminution du poids des gonades était plus considérable chez les poissons femelles. Les deux espèces les plus fréquemment utilisées, à l'échelle nationale, étaient le meunier noir et le meunier rouge, et les types de réponse pour ces deux espèces concordaient avec l'effet moyen national (perturbation métabolique et enrichissement en substances nutritives). Des réponses similaires ont été observées autant pour le traitement aux boues activées que pour l'utilisation d'étangs de stabilisation aérobie, et autant pour le procédé kraft que pour le procédé thermomécanique; elles concordaient toutes avec le type moyen de

réponse à l'échelle nationale. Il y avait des indications de réponses différentes selon les principaux types d'habitat, mais, étant donné que la plupart des études sur les poissons étaient réalisées dans des habitats fluviaux, une évaluation plus détaillée de l'influence du type d'habitat nécessitera des données supplémentaires.

Les résultats méta-analytiques des études sur les communautés d'invertébrés benthiques et les populations de poissons ont été quantitativement intégrés à l'aide d'une ordination d'échelonnement multidimensionnel. Le type global de réponse a montré l'existence d'une bonne concordance entre les résultats des deux études. La perturbation métabolique et l'enrichissement en substances nutritives, observés chez les poissons, chevauchaient les résultats de l'eutrophisation constatés chez le benthos. Les résultats ont aussi montré que l'enrichissement total en substances nutritives des poissons (sans perturbation métabolique mesurable) n'était peut-être pas décelé avant que le benthos atteigne un état d'eutrophisation prononcée. Qui plus est, les résultats obtenus pour les poissons ont indiqué l'existence d'effets de perturbation métabolique, même lorsque l'effet général en était un d'eutrophisation. Par conséquent, les deux éléments principaux du SEE sur le terrain se complètent mutuellement en révélant des aspects différents des effets associés à une exposition aux effluents.

Les données sur la toxicité sublétale ont montré une amélioration, à l'échelle nationale, de la qualité des effluents des fabriques canadiennes de pâtes et papiers entre le premier et le deuxième cycle. Cette tendance nationale à l'amélioration était particulièrement évidente dans les niveaux de toxicité pour *Ceriodaphnia*, la tête-de-boule, *Selenastrum*, et les capucettes béryl. Les résultats des essais sur les échinodermes et *Champia parvula* ont montré des tendances moins frappantes en ce qui a trait à la diminution de la toxicité pendant le deuxième cycle. Ces diminutions de la toxicité des effluents entre le premier et le deuxième cycle sont largement attribuées à la mise en place de systèmes de traitement secondaire dans les fabriques du pays.

Les données du deuxième cycle ont été recueillies deux ans seulement après l'installation dans bon nombre de fabriques d'un système de traitement secondaire, et elles ont déjà commencé à montrer les premiers signes d'un rétablissement au voisinage de quelques fabriques ainsi qu'une diminution de l'occurrence d'effets sublétaux chez les communautés d'invertébrés benthiques. En résumé, la qualité des effluents s'est fortement améliorée depuis la promulgation, en 1992, du REFPP, mais des effets sur les invertébrés benthiques et le poisson sont encore constatés. Il faut poursuivre les études de suivi pour évaluer l'étendue spatiale et l'importance écologique des effets observés, et Environnement Canada continuera de collaborer avec l'industrie et d'autres intervenants afin de mieux comprendre leur importance.

Table of Contents

1. INTRODUCTION 1

1.1 Objectives of the Report 1
1.2 The National EEM Program for Pulp and Paper 1
1.3 Rationale for this Study 3
1.4 Outline of this Report 5

2. HISTORICAL REVIEW OF PULP AND PAPER MILL EFFLUENTS 6

2.1 Overview of the Canadian Pulp and Paper Industry 6
2.2 Impact on the Environment 8
2.2.1 Impact on Fish 9
2.2.2 Impact on Benthic Invertebrates 11

3. GENERAL METHODS 14

3.1 Overall Data Preparation and Analysis 14
3.2 Synthetic Analyses 16

4. BENTHIC INVERTEBRATE COMMUNITY SURVEY 19

4.1 Introduction 19
4.2 Comparison with Cycle 1 Results 19
4.3 Data Extraction and Preparation for Analysis 20
4.4 National Trends 24
4.4.1 Overall Response Patterns 24
4.4.2 Habitat 32
4.4.3 Effluent Treatment and Mill Process Types 37
4.5 Evaluation of Selected Benthic Survey Program Elements 44
4.5.1 Relative Sensitivity of Endpoints 44
4.5.2 Study Design 47
4.5.3 Effect Size for Calculating Required Sampling Effort 49
4.6 Conclusions 53

5. FISH SURVEY 54

5.1 Introduction 54
5.2 Improvement in the Fish population survey — Cycle 1 versus Cycle 2 56
5.3 Data Extraction and Preparation for Analysis 59
5.3.1 Data Extraction and Preparation 59
5.3.2 Discussion of Data Entry Errors 60
5.4 National Trends — ANCOVA 61
5.4.1 Results from ANCOVA 61
5.4.2 Interpretation of ANCOVA Results 66
5.5 National Trends — Meta-analyses 69
5.5.1 Results from Meta-analyses 69
5.5.1.1 Gender 70
5.5.1.2 Species 70
5.5.1.3 Mill Process Type 72
5.5.1.4 Treatment Type 73
5.5.1.5 Habitat 74
5.5.2 Interpretation of Meta-analyses 75
5.6 Evaluation of Selected Fish population survey Program Elements 77
5.6.1 Effect Size for Calculating Required Sampling Effort 77
5.6.2 Evaluation of Numbers of Significant Differences for Each Endpoint 78
5.7 Conclusions and Recommendations 79

6. OVERALL NATIONAL RESPONSE PATTERNS FOR BENTHIC INVERTEBRATE COMMUNITY AND FISH POPULATION SURVEYS 82

6.1 BENTHIC INVERTEBRATE COMMUNITY SURVEY 83
6.2 FISH POPULATION SURVEY 85
6.3 COMBINED BENTHIC INVERTEBRATE AND FISH RESPONSES 88

7. INTERPRETATION OF SUBLETHAL TOXICITY DATA FOR CYCLE 2 OF THE PULP AND PAPER EEM PROGRAM 92

7.1 INTRODUCTION 92
7.2 SUBLETHAL TOXICITY TEST METHODS 92
7.3 MONITORING CHANGES IN EFFLUENT QUALITY BETWEEN CYCLES 93
7.4 MONITORING CHANGES WITH TREATMENT OF EFFLUENT 98
7.5 CONCLUSIONS 101

8. SUMMARY AND CONCLUSIONS 102

8.1 HISTORICAL CONTEXT 102
8.2 COMPONENTS OF THE NATIONAL EEM PROGRAM 103
8.2.1 Field Study Components 103
8.2.1.1 Benthic Invertebrate Community Survey 104
8.2.1.2 Fish Population Survey 104
8.2.2 Sublethal Toxicity Testing Component 105
8.3 CYCLE 2 RESULTS 105
8.3.1 Benthic Invertebrate Community Survey 105
8.3.1.1 Habitat Type 106
8.3.1.2 Effluent Treatment Type 107
8.3.1.3 Mill Process Type 107
8.3.2 Fish Population Survey 108
8.3.2.1 Gender 109
8.3.2.2 Fish Species 109
8.3.2.3 Mill Process Type 109
8.3.2.4 Effluent Treatment Type 109
8.3.2.5 Habitat 110
8.3.3 National Response Patterns for Combined Benthic Invertebrate Community and Fish Population Surveys 110
8.3.4 Sublethal Toxicity Testing 110
8.4 CONCLUSIONS AND RECOMMENDATIONS 111
8.4.1 Endpoints 111
8.4.1.1 Benthic Invertebrate Community Survey 111
8.4.1.2 Fish Population Survey 112
8.4.2 Study Design 112
8.4.2.1 Benthic Invertebrate Community Survey 112
8.4.2.2 Fish Population Survey 113
8.4.3 Effect Size for Calculating Required Sampling Effort 113
8.4.3.1 Benthic Invertebrate Community Survey 113
8.4.3.2 Fish Population Survey 114
8.4.4 Data Presentation 114
8.4.5 Cycle 2 Results 114
8.4.5.1 Benthic Invertebrate Community Survey 114
8.4.5.2 Fish Population Survey 114
8.4.5.3 Sublethal Toxicity Testing 115
8.4.6 Ongoing Analyses and Interpretation 115
8.5 OVERALL CONCLUSIONS 116

REFERENCES 117

LIST OF TABLES

Table 4-1: Greatest significant increases in abundance 31
Table 4-2: Greatest significant decreases in abundance 31
Table 4-3: Greatest significant increases in taxon richness 32
Table 4-4: Greatest significant decreases in taxon richness 32
Table 4-5: Number of mills in different habitat categories 33
Table 4-6: Number of mills in different effluent treatment and mill process categories 38
Table 5-1: The number of mills that encountered various problems in the Cycle 1 fish population survey within the freshwater receiving environment (from Environment Canada, 1997c) .. 56
Table 5-2: Species used (in order of frequency in Cycle 1 and Cycle 2) at freshwater and marine/estuarine sites 57
Table 5-3: Endpoints used to determine effects in EEM, as determined by statistically significant differences between exposure and reference areas 59
Table 5-4: Data entry and analysis problems found in the review of the electronic data submitted for Cycle 2 EEM reports (from Kilgour, 2001) 60
Table 5-5: Sites with the largest percentage of (A) decreases and (B) increases in gonad weight in female finfish sampled during Cycle 2 (percent difference is given as exposure minus reference mean, expressed as a percentage of the reference mean). For comparison, corresponding changes in male gonad weight, female condition factor, and female liver weight are also given. 65
Table 5-6: Percent change in condition, gonad, and liver weight and the corresponding percentile 66
Table 5-7: List of all species that were included in the meta-analysis 71
Table 7-1: Sublethal toxicity tests for EEM 93

List of Figures

Figure 2-1: Schematic diagram illustrating the impacts of nutrients and toxicity, effluent treatment types, and receiving environment dilution on community structure and individual growth/survival rates of aquatic biota exposed to pulp mill effluent (obtained from Lowell et al., 2000) 12
Figure 3-1: Simplified flow chart of main steps in data preparation and analysis 14
Figure 3-2: Meta-analysis summary figure 18
Figure 4-1: Percentage of interpretative reports reporting an improvement or worsening in benthic conditions in exposure areas relative to reference areas in Cycle 2 (of mills reporting a change) 20
Figure 4-2: Comparison of electronic versus interpretative report data: Percentage of mills where statistically significant effects were measured for one, both, or neither of the two endpoints (abundance, taxon richness) 23
Figure 4-3: Predicted response patterns (see text; modified from Lowell et al., 2000; see also Culp et al., 2000a) 24
Figure 4-4: Control/Impact mill grand means 26
Figure 4-5: Control/Impact + Gradient mill grand means 28
Figure 4-6: Abundance measured effect size distribution 29
Figure 4-7: Taxon richness measured effect size distribution 30
Figure 4-8: Evenness measured effect size distribution 30
Figure 4-9: Bray-Curtis measured effect size distribution 31
Figure 4-10: Abundance and taxon richness by habitat (Control/Impact mills) 33
Figure 4-11: Abundance and taxon richness by habitat (Control/Impact + Gradient mills) 34
Figure 4-12: Evenness by habitat (Control/Impact mills) 35
Figure 4-13: Evenness by habitat (Control/Impact + Gradient mills) 36
Figure 4-14: Bray-Curtis by habitat (Control/Impact mills) 37
Figure 4-15: Abundance and taxon richness by effluent treatment (Control/Impact mills) 39
Figure 4-16: Evenness by effluent treatment (Control/Impact mills) 40
Figure 4-17: Bray-Curtis by effluent treatment (Control/Impact mills) 41
Figure 4-18: Abundance and taxon richness by mill process (Control/Impact mills). Kraft = kraft mills; TMP = thermomechanical pulp mills. 42
Figure 4-19: Evenness by mill process (Control/Impact mills). Kraft = kraft mills; TMP = thermomechanical pulp mills. 43
Figure 4-20: Bray-Curtis by mill process (Control/Impact mills). Kraft = kraft mills; TMP = thermomechanical pulp mills. 44
Figure 4-21. Percentage of Control/Impact mills showing significant exposure versus reference differences 45
Figure 4-22: Evenness versus diversity comparison (Control/Impact marine/estuarine mills) 47
Figure 4-23: Percentage of mills showing significant exposure versus reference differences for zero, one, or a combination of three endpoints (abundance, taxon richness, evenness) for Control/Impact and Gradient mills 48
Figure 4-24: Measured effect size (ES) and "normal range of variability (2SD)" (relative to reference areas) 51
Figure 4-25: Percentage of mills where measured effect size exceeded "normal range of variability (2SD)" 52
Figure 5-1: Comparison of female (black box) and male responses (hatched box) for condition factor (A) and relative changes in liver weight (B) and gonad weight (C) 61
Figure 5-2: Percent changes in condition factor (A) and relative changes in liver weight (B) and gonad weight (C). Data sets showing interactions were excluded from this analysis. 63

Figure 5-3: Cycle 2 fish responses for condition (black, n = 140), liver weight (shaded, n = 155), and gonad weight (clear boxes, n = 154). Values are given as the relative difference between the exposure versus reference area means, expressed as a percentage of the reference area mean 67
Figure 5-4: Detransformed percent changes in effect size relative to the reference mean 69
Figure 5-5: Effect of gender on the five key fish endpoints. M = male, F = female. Error bars represent 95% confidence intervals 70
Figure 5-6: Effect of species on the five key fish endpoints. WS = white sucker; LS = longnose sucker; TP = trout-perch; YP = yellow perch. Error bars represent 95% confidence intervals 72
Figure 5-7: Effect of mill process type on the five key fish endpoints. Kraft = kraft mills; TMP = thermomechanical pulp mills. Error bars represent 95% confidence intervals 73
Figure 5-8: Effect of treatment type on the five key fish endpoints. ASB = aerated stabilization basin; AST = activated sludge treatment. Error bars represent 95% confidence intervals ... 74
Figure 5-9: Effect of habitat on the five key fish endpoints. R = riverine; L = lacustrine; M/E = marine/estuarine. Error bars represent 95% confidence intervals 75
Figure 5-10: Percentage of fish comparisons in Cycle 2 where the measured effect size exceeded two standard deviations (2SD), 25%, or 10% 78
Figure 5-11: Types of statistically significant endpoints observed at each mill site using data extracted from the mill interpretative reports (filled boxes, n = 53) and the electronic data (hatched boxes, n = 62). Results are presented as the percentage of times that mills found specific combinations of these responses — condition factor (k), gonad weight, and liver weight 79
Figure 6-1: Overall national response pattern for benthic invertebrates: abundance and taxon richness endpoints 83
Figure 6-2: Overall national response pattern for benthic invertebrates: all four endpoints 84
Figure 6-3: Overall national response pattern for fish: gonad weight, condition factor, and liver weight endpoints 86
Figure 6-4: Overall national response pattern for fish: all five endpoints 87
Figure 6-5: Overall national response pattern for benthic invertebrates and fish combined: two benthic and three fish endpoints 89
Figure 6-6: Overall national response pattern for benthic invertebrates and fish combined: all benthic and fish endpoints 90
Figure 7-1: Comparison of sublethal toxicity to freshwater species in Cycles 1 and 2 of EEM ... 95
Figure 7-2: Comparison of sublethal toxicity to marine species in Cycles 1 and 2 of EEM 96
Figure 7-3: Percentages of sublethal tests showing no effect at highest concentration 97
Figure 7-4: Sublethal toxicity of Ontario mill effluents before secondary treatment (Cycle 1 of EEM) and after (Cycle 2) 99
Figure 7-5: Comparison of sublethal toxicity of secondary-treated effluent from freshwater mills in British Columbia in Cycles 1 and 2 of EEM 100

1. Introduction

1.1 Objectives of the Report

The purpose of this report is to present to the scientific community the results of a national assessment of environmental effects monitoring (EEM) data, which was primarily designed to determine the type and magnitude of effects of pulp and paper mill effluent on fish and benthic invertebrate communities. Environment Canada scientists guided the data analysis through the development of a series of key questions or priority areas of focus, such as the following:

- What are the types and magnitude of "effects" of pulp and paper mill effluent on fish and benthic invertebrate communities, and how are these effects influenced by factors such as mill process, treatment type, habitat type, study design (in the case of the benthic invertebrate community survey), and gender differences (specific to fish)?

- How do the various fish and benthic invertebrate community EEM effects endpoints relate to each other? Do these endpoints provide complementary or contradictory information?

- What is the relative sensitivity of the various effects endpoints in determining the presence of effects?

- What are the "critical effect sizes" for each EEM endpoint (based on EEM data collected to date) that can be used to guide the development of consistent national sampling designs?

These questions are fundamental to evaluating the effectiveness of EEM approaches in assessing the effects of pulp and paper mill effluent on aquatic environments and providing the necessary information to understand the ecological importance of identified EEM effects — in other words, to help determine when an effect in EEM is considered an environmental problem.

In addition to the analysis of the priority areas mentioned above, sublethal toxicity data on pulp mill effluent were analyzed to assess overall changes in effluent sublethal toxicity before and after mills implemented secondary treatment technologies. Supporting elements of the National EEM Program (i.e., dioxins and furans, tainting, water and sediment quality) are not the subject of this report, but will undergo analysis at a future date.

1.2 The National EEM Program for Pulp and Paper

In 1992, the *Pulp and Paper Effluent Regulations* (PPER) under the *Fisheries Act* replaced a 1971 pulp and paper regulation. The PPER set discharge limits for total suspended solids (TSS) and biochemical oxygen demand (BOD). As well, they set a requirement that all discharged effluents should be non-acutely lethal to rainbow trout in 100% effluent. Compliance with the PPER entailed major changes in the way effluents were treated by the industry, resulting in most cases in the installation of secondary (biological) treatment plants. Although it was acknowledged that more stringent discharge limits would improve environmental protection, it was also recognized that these measures alone might not ensure adequate protection of the

aquatic ecosystem at every site. Consequently, the 1992 Regulations included the requirement for an EEM program.

The National EEM Program requires Canada's pulp and paper mills to conduct studies on their receiving environments in order to assess and monitor effects potentially caused by their effluent. EEM's site-specific nature calls for iterative evaluations of the potential effects of effluent on fish, fish habitat, and the use of fisheries resources. The program is structured in a three-year sequence of monitoring and interpretation phases known as "cycles." At the beginning of each cycle, each mill is expected to develop a site-specific monitoring program in collaboration with regional Environment Canada officials. At the end of each cycle, mills are required to submit an interpretative report summarizing their field work and interpreting their results. The structure of the EEM program ensures a certain level of national consistency in the way in which mills monitor the effects of their effluent on the environment.

The EEM program is unique, in that it provides a long-term evaluation of effluent impacts on the aquatic environment. To be effective, a minimum of several cycles (15–20 years) is required to allow an assessment of the condition of the receiving environment and a true understanding of the significance of any impacts. Moreover, EEM provides quantitative and qualitative answers to clear, well-defined questions by measuring responses of organisms and populations to the presence of effluents in exposure and reference areas (Hodson et al., 1996). The EEM model has been adapted to other industries interested in monitoring their environmental impacts over the longer term. The program uses a tiered approach to monitoring, with initial studies carried out to characterize and assess the condition of the receiving environment, followed by targeted studies once effects are found or reduced monitoring where effects are not occurring. This structure minimizes costs but allows the program to remain scientifically defensible and flexible, so that new or improved monitoring techniques can be incorporated.

An EEM study includes the following components:

- a fish population survey to assess the health of fish;
- a benthic invertebrate community survey to assess fish habitat;
- a study of dioxins and furans in edible fish tissue and a tainting study as assessments of the usability of fisheries resources;
- sublethal toxicity testing to assess effluent quality; and
- supporting water and sediment quality variables to aid in the interpretation of biological data.

The fish population survey and benthic community survey utilize a select group of "effects" endpoints that primarily drive future monitoring but also contribute to understanding the real nature of impacts (at the site-specific level) from pulp and paper effluent discharges. In EEM, an "effect" is a statistically significant response in at least one of the select endpoints in comparisons between biological samples taken near a mill (exposure area) and samples taken from a reference area. The reference area is a sampling area as similar as possible in all aspects to the exposure area (e.g., same habitat, water quality features, hydrological features, etc.), but without the presence of mill effluent. Unlike the biological field data, the supporting variables and sublethal toxicity data were not used to determine whether or not environmental "effects"

are occurring at mill sites. Rather, they are meant to provide further information on parameters that may or may not have an influence on any observed effects.

The EEM effects endpoints used in the fish and benthic invertebrate community surveys are listed below:

Fish Population Survey Endpoints:

- Gonad weight
- Condition factor
- Liver weight
- Weight at age
- Age

Benthic Invertebrate Community Survey Endpoints:

- Total abundance
- Taxon richness
- Bray-Curtis index of dissimilarity
- Simpson's diversity (or evenness)

The fish population survey endpoints stem from a framework originally developed to describe the responses of fish populations to anthropogenic stress and are meant to assist in designing studies aimed at identifying stressors associated with changes in the performance of fish populations (Gibbons and Munkittrick, 1994). This framework was further developed, and population parameters were grouped according to summary categories of age structure, energy expenditure (gonad weight), and energy storage (condition and liver weight) (Environment Canada, 1997c). Weight at age is a measure of the growth rate described by the relationship of size (weight) to age. Generally, over the life of a fish, the rate of increase in weight declines as fish approach the limit of their life span (Environment Canada, 1998b).

The EEM benthic invertebrate community endpoints are similar to those used in many benthic surveys and are measured to evaluate the benthic community structure and its functioning. These endpoints are used in the EEM program to determine the magnitude and spatial extent of habitat degradation. Overall invertebrate abundance can often provide additional information on the energy available to the next trophic level (i.e., fish), while community descriptors, such as taxon richness and diversity, can provide information on structural changes over time. The Bray-Curtis index provides quantitative information on changes in the composition of taxa between exposure and reference areas that could potentially go unnoticed by looking at abundance, richness, and diversity alone. Later in this report, the utility of using evenness instead of diversity is explored.

1.3 Rationale for this Study

The foundation for this national analysis of EEM data is the fish and benthic community surveys using the endpoints identified above to answer some of the key questions related to the potential presence and type of effects as the result of exposure to pulp and paper mill effluent. As mentioned previously, the pulp and paper EEM studies are conducted every three years, and each round of monitoring and interpretation is known as a "cycle." Since 1992, the pulp and paper industry has completed two cycles of monitoring and reporting. The industry is currently in the middle of Cycle 3. Although this assessment is based primarily on Cycle 2 data, some comparisons are made with Cycle 1. The reasons behind this are explained below. It should also be noted that at the time Cycle 2 data was collected, many mills had only recently installed secondary treatment.

Cycle 1 was primarily used as a baseline to gain a better understanding of the variability of the field measurements. Although the methodologies developed for the EEM program were founded on good science and tested successfully in investigations outside of EEM, Cycle 1 was the first national-scale monitoring effort using the EEM monitoring approaches in Canada. As such, it was not surprising that a number of monitoring problems were identified after the end of the first EEM cycle. This led to an extensive government–industry review of Cycle 1 to identify the specific monitoring problems and to provide recommendations for improvements for the following cycles. "Decision trees" were developed to help guide mills, consultants, and Environment Canada's regional co-ordinators in determining site-specific study designs (e.g., fish species selection and selection of the reference area). A portion of the Cycle 1 review also included modifying the Technical Guidance Document (Environment Canada, 1998b) to provide mills and consultants with more detailed guidance on monitoring on a site-specific basis.

Many of the problems identified following Cycle 1 were associated with field monitoring approaches. Moreover, issues related to generating good quality data using certain sublethal toxicity test procedures were also identified. Some of the problem areas identified for the field approaches included:

- poor selection of fish and benthic invertebrate community reference sites (for comparison against exposure sites);
- inability to catch sufficient fish for the fish population survey;
- difficulties in controlling for "confounding" factors when interpreting data (e.g., presence of natural or anthropogenic sources other than the mill that could cause effects); and
- mixing of exposure and reference fish.

Overall, the data generated for the laboratory element of EEM in Cycle 1 (sublethal toxicity testing of effluent) were of relatively good quality, whereas the field data for the benthic invertebrate community and fish population surveys were considered to be of "poorer" quality. These differences in data quality are accounted for by the very nature of the different types of monitoring. Sublethal toxicity testing makes use of uniform methodologies, whereas the monitoring approaches used when conducting field work require a good level of "flexibility" to account for variations. In other words, the ability to generate nationally consistent, good quality data for the fish population survey and benthic invertebrate community survey is generally more difficult because of the complexity of natural environments compared with the more artificial, controlled nature of laboratory tests. The Cycle 1 review provided recommendations to deal with these issues (including more detailed monitoring guidance), leading to an improvement of Cycle 2 study designs and, consequently, more consistently good quality data. For example, in Cycle 1, only 8% of mills (9 of 115 studies) successfully caught the required minimum number of fish (at least 20 fish of each sex of two fish species) in the exposure area and reference area (Environment Canada, 1997c) for the fish population survey. In Cycle 2, however, over 56% of mills (64 of 114 interpretative reports) that attempted a standard fish population survey succeeded in capturing 20 males and 20 females of two fish species. Consequently, the focus of this national assessment is placed on Cycle 2 data, with some comparisons with Cycle 1 data.

1.4 Outline of this Report

This report is divided into the following chapters:

- Chapter 2 presents a historical review of the effects of pulp and paper mill effluents with respect to the EEM endpoints and is later used as a basis for comparison with what the data are currently showing.
- Chapter 3 outlines the methodologies that were employed for data analysis.
- Chapters 4 and 5 describe the results for the benthic invertebrate community and fish surveys, respectively.
- Chapter 6 shows the results of a combined analysis of fish population survey and benthic invertebrate community survey endpoints.
- Chapter 7 highlights the results of the sublethal toxicity data analysis.
- Chapter 8 provides an overall summary of report conclusions, recommendations, and next steps for analysis.

2. Historical Review of Pulp and Paper Mill Effluents

2.1 Overview of the Canadian Pulp and Paper Industry

Over the past 40 years, the pulp and paper industry has greatly influenced Canada's socio-economic and environmental systems. The industry counts over 100 pulp and paper mills across the country and has been an important source of employment in many Canadian communities. Canada is one of the world's largest producers and exporters of pulp and paper and exports 80% of its pulp and paper products to approximately 100 countries around the world (FPAC, 2001).

The manufacturing of writing, printing, and wrapping papers in Canada dates back to the early 1800s. The Canadian pulp and paper industry survived the Great Depression in the 1930s, increasing international competition, and environmental concerns through the 1970s and is currently a multibillion-dollar business, employing several thousand people (FPAC, 2001). The pulp and paper industry in Canada has had a great influence on the Canadian economy and continues to influence it on the basis of both domestic revenues and exports. The industry is very diverse, and the 134 mills currently conducting EEM are divided into approximately 30 different categories based on process type. The largest of these is the kraft mill category, with 39 mills, followed by thermomechanical, with 24 mills.

Pulp and paper products are made from three major types of fibres: wood, recycled fibres, and non-wood fibres (e.g., cotton, hemp, etc.), however most pulp and paper products are wood-based. Kraft mills use a variety of chemicals in their pulping processes and have been the source of much research attention over the past two decades. In the kraft process, wood chips are turned into pulp by chemically dissolving the lignin binding the wood fibres. These pulps are generally stronger than mechanical pulps. Kraft pulp is used for products such as writing papers, paper bags, and cardboard. Other pulping procedures involving primarily chemical processes include those used at soda and sulphite mills. Mechanical pulping involves shredding or grinding wood chips and usually results in high-yield pulps; that is, a very high percentage of the original wood components is retained in the final product. Mechanical pulps are most commonly used for newsprint and catalogue paper (Biermann, 1996). Finally, semi-chemical pulping combines chemical and mechanical methods, where wood chips are partially digested in a weak chemical solution followed by mechanical refining for fibre separation (USEPA, 1998). Semi-chemical pulp is commonly used for manufacturing corrugated cardboard.

Effluent treatment processes include primary, secondary, and in a few cases tertiary treatments. In primary treatment, solids are removed from water by allowing them to settle out. This method became widely used in the 1950s (Biermann, 1996). Today, most mills pass their effluents through at least a secondary treatment process as well, which involves reacting the effluent (after primary treatment) with oxygen and microorganisms to remove oxygen-consuming materials. This process reduces biochemical oxygen demand (BOD) by 80–90% and decreases the effluent toxicity significantly (Biermann, 1996). Secondary treatment can be carried out using an aerated stabilization basin, an activated sludge process, an oxidation basin, or rotating biological filters, with the first two methods being most commonly used in Canada. Tertiary treatments are the most advanced types of treatments; they follow secondary treatment and are often considered for

the removal of colour (Biermann, 1996). By the mid-1990s, most mills in Canada had adopted primary and secondary treatment processes.

The diverse nature of the pulp and paper industry and its processes is an indication of its potential for releasing a wide range of compounds into the aquatic environment. A substantial amount of research has been done on pulp and paper effluents, implicating fibre and suspended solids, colour and turbidity, and organic and nutrient enrichment loads as the three conventional pollutant factors with adverse environmental impacts (Owens, 1991). More specifically, BOD, TSS, and acute toxicity were among the major issues identified as having negative impacts on aquatic ecosystems. In the late 1980's, some of the chlorinated organic compounds synthesized during pulp bleaching were characterized as persistent and toxic at low concentrations. Dioxins and furans released in pulp and paper effluent were also a concern for communities living downstream from mills, particularly those with high fish consumption. In fact, in 1990, pulp and paper mills were identified as the main industrial source of dioxins and furans in Canada (Environment Canada, 1990). These compounds, which are highly persistent in the environment, have a strong affinity for sediments, and are very bioaccumulative, became regulated under the *Pulp and Paper Chlorinated Dioxins and Furans Regulations* of the *Canadian Environmental Protection Act* (CEPA) in 1992.

In 1992, the federal government passed the Pulp and Paper Regulatory Framework consisting of the *Pulp and Paper Effluent Regulations* (PPER) under the *Fisheries Act* and two regulations under the *Canadian Environmental Protection Act*. The PPER revoked and replaced an earlier set of regulations passed in 1971. A major impetus for passage of the new regulations was to ensure that all mills were subject to regulatory requirements. A major deficiency of the 1971 regulations, was that they were only legally binding on new mills built after the regulation's November 2, 1971 promulgation date. The PPER established discharge limits for BOD, TSS and prohibit the discharge of effluent that is acutely lethal to rainbow trout. The CEPA regulations were designed to ensure that chlorinated dioxins and furans were not formed during pulp bleaching. The *Pulp and Paper Mill Effluent Chlorinated Dioxins and Furans Regulations* prohibit the release of measurable concentration of the 2,3,7,8 chlorinated dioxin and furan in effluent from mills that use chlorine or chlorine dioxide to bleach pulp. The *Pulp and Paper Mill Defoamer and Wood Chip Regulations*, impose quality requirements for defoamers used in chlorine bleaching processes and prohibit the manufacture of pulp from wood chips treated with polychlorinated phenols (OECD, 1999).

The implementation of these new regulations led to major capital investments by the industry in the effluent treatment processes at pulp and paper mills, which significantly improved the effluent quality. However, it was recognized that these measures alone might not ensure adequate protection of the aquatic ecosystem at every site. It was also recognized that the assessment of the environmental impact of pulp and paper mill effluents was a difficult task, as a result of the drastic variability in the chemical composition of pulp and paper mill effluents as well as the complexity and natural variability of biological systems. Consequently, the PPER included the addition of a requirement for an EEM program in order to provide an overview of the state of receiving environments around pulp and paper mills across Canada, to attempt to demonstrate whether or not the major improvements in effluent quality resulting from improvements in treatment processes were associated with similar improvements in the receiving

environment, and to evaluate the long-term effects of pulp and paper mill effluents on Canadian aquatic ecosystems.

The EEM program is the first legislated program of its kind and was introduced as a tool to evaluate the long-term effects of pulp and paper mill effluents on Canadian aquatic ecosystems. Its very nature goes beyond the traditional end-of-pipe measurements and requires the collection of data at a national scale so as to provide an overview of the state of receiving environments across Canada and to aid in identifying site-specific impacts. EEM data collected were used to evaluate the overall effects of the discharge of pulp and paper mill effluents on receiving waters and to show the presence of any discernible trends. This information, taken together with social, economic, and technical considerations, can eventually be used as a tool to evaluate the adequacy of the regulations in protecting fish populations, their habitat, and the human use of fisheries resources and to ensure proper management, conservation, and protection of aquatic environments.

The implementation of these new regulations in combination with increasing environmental pressures led to an important improvement in the industry's treatment processes so as to greatly reduce the impact of their effluents on the receiving environment. For instance, over the 1990s, dioxins and furans were largely eliminated from pulp and paper mill effluents by replacing elemental chlorine with less harmful substitutes in the pulp bleaching process. In fact, between 1990 and 1997, almost a 99% reduction in dioxins and furans was achieved compared with base year 1990 following the implementation of the *Pulp and Paper Effluent Regulations* (Environment Canada, 1999b).

2.2 Impact on the Environment

As previously mentioned, pulp and paper mill effluents have been known to release a number of pollutants with adverse environmental effects. These include BOD, TSS, colour and turbidity, organic and nutrient enrichment loads, chlorinated organics, and dioxins and furans, among others. Although the industry has successfully mediated some of these issues, concerns remain about the long-term impacts of effluents on the aquatic environment. The chemical composition of each effluent varies drastically, making the assessment of their environmental impact a difficult task. Moreover, the complexity and natural variability of biological systems further make it difficult to provide definitive answers with regard to the environmental impacts of effluents (Kovacs et al., 1995). However, the conceptual approaches used in the EEM program, based on site-specific design and applying standardized methods, have clearly shown that knowledge on the environmental effects of effluents can be gained. For the first time, this national assessment is providing answers on the effects of effluent based on geographically extensive scientific data collected in the receiving environment.

The 1980s brought world-wide attention on bleached pulp mill effluents and their environmental impacts, as some research shifted in focus from the more conventional parameters, such as BOD, TSS, and nutrients, towards chemical toxicity (Lehtinen, 1996). The formation of organic compounds with varying molecular size was suspected of being the specific cause of a number of biological effects, prompting researchers to study relationships between exposure and effects (Lehtinen, 1996). Studies were also showing that pulp mill effluents were creating imbalances in

fish community structures and abnormal size distribution and age structure (Adams et al., 1992). Potential impacts on "the reproductive performance of fish, including circulating sex steroid levels, gonad size, age to maturity, and secondary sex characteristics" (McMaster et al., 1991; Munkittrick et al., 1991, 2000), were also being observed. With respect to benthic invertebrate communities, pulp and paper mill effluents were showing impacts on their downstream structure and functioning, with communities often exhibiting fewer species, lower densities, and altered taxonomic composition relative to those at a reference site (Kelso, 1977; VanderWal, 1977; Mayack and Waterhouse, 1983; Scrimgeour, 1989).

Changes implemented as a result of the 1992 regulatory package led to dramatic improvements in the chemical composition and toxicity of the final effluent, significantly reducing the potential threat to the environment (Bothwell, 1992; Servos et al., 1996; Dubé and McLatchy, 2000). In the late 1990s, studies were beginning to document the early signs of recovery in fish reproductive parameters at a number of mills that had modernized their processes (Munkittrick et al., 1997). Advanced pulp mill effluent treatment also resulted in significantly reduced incidences of toxic effects on or pronounced eutrophication in benthic invertebrate communities (Lowell et al., 1996, 2000; Felder et al., 1998; Chambers et al., 2000b; Culp et al., 2000a,b; Lowell and Culp, 2002).

2.2.1 Impact on Fish

Studies in North America in the late 1980s found a wide variety of responses in fish exposed to primary-treated effluent from a bleached kraft pulp mill (McMaster et al., 1991; Munkittrick et al., 1991). Additional studies followed the responses of fish at that site to installation of secondary effluent treatment and changes in pulping process (reviewed in Munkittrick et al., 1997, 1998). Expansion of those studies to other large pulp mills in Ontario suggested that there was a potential for effluents from large pulp mills with secondary treatment to be associated with impacts on the reproductive development of fish (Munkittrick et al., 1994). The potential of some effluents to cause reproductive impairment in fish was confirmed in laboratory studies (Kovacs et al., 1995).

The introduction of stricter regulations on pulp mill effluents in 1992 led to significant improvements in mill process and treatment, reducing the production and release of harmful substances into the aquatic environment. Among these advances was the elimination of acute toxicity in whole effluent and a reduction in the production and release of persistent organochlorines. Despite these advances, effects on wild fish were still being observed. These included changes in fish reproduction, condition, size at age, liver weight, and age distributions, among others.

In the early 1990s, the environmental fate and effects of pulp and paper mill effluents became a very active area of research, and studies consistently showed significant physiological changes in wild fish downstream of Canadian pulp and paper mills (Munkittrick et al., 1997). Moreover, researchers were observing that, despite mills' adherence to the regulations governing dioxins, oxygen demand, and acute lethality, some of their effluents could still be associated with subtle changes in the reproduction of fish. A number of compounds in pulp mill effluent have been identified as causing changes in reproductive parameters; however, those compounds responsible

for the physiological impacts remain a mystery (Munkittrick et al., 1998). Nevertheless, natural compounds, additives, and compounds originating in condensates, black liquor, and waste treatment have been shown to affect physiological functions in fish (Munkittrick et al., 1998). These physiological changes appear to be reversible and are causing scientists to reexamine the role of persistent organochlorines with respect to the health of fish near pulp mill discharges (Munkittrick et al., 1998).

The Munkittrick et al. (2000) Moose River Basin studies (from 1993 to 1998) have shown that resident fish do respond to changes in their environment and that they provide an integrated response to stressors in the ecosystem. The collection of fish at many sites over several years makes it possible to follow trends in data over time (Munkittrick et al., 2000), which is the primary goal of the EEM program. At Kapuskasing and Smooth Rock Falls, the pulp mills underwent extensive changes and modernization between 1991 and 1998, and responses of fish over this period of modernization were used to examine the relative contributions of the mill to fish performance (Munkittrick et al., 2000). At the Kapuskasing mill, changes were initiated in late 1991, and, by the fall of 1993, its processes were switched over to thermomechanical pulp. That fall, Munkittrick et al. (2000) noted a marked increase in the liver size of white suckers, but following the completion of mill process changes, no reductions in gonad size were observed. In 1994, fish collected at a site within 2 km of the discharge showed much larger livers than downstream fish, but again gonad size was not reduced. Following the addition of secondary treatment in the fall of 1994, researchers found a substantial decrease in liver size, although these were still larger than at reference sites, while condition factor also showed a gradual decline back to reference levels. In contrast, the recovery in gonad size occurred following the mill process changes, but before the addition of secondary treatment. Fish collected at Smooth Rock Falls showed similar responses following process alterations and the addition of secondary treatment in 1994. Improvements in fish performance at this site were also attributed to changes in the pulping process to reduce black liquor losses. These Moose River Basin results are supported by a number of other studies showing improvements in fish responses following mill process alterations (McMaster et al., 1991; Munkittrick et al., 1991; Munkittrick and McMaster, 2000).

The review of EEM data has shown that there are a variety of types of responses of fish populations (these are explained in more detail in Chapter 5). Generally, fish life history characteristics (size to maturity, growth rate, fecundity) will respond in parallel to changing environmental conditions or the presence of stressors (Munkittrick et al., 2000). The most common responses to pulp mill effluent exposure have been eutrophication (an increased availability of food resources associated with nutrient input in the effluent), metabolic disruption (associated with a decreased utilization of available food resources, presumably associated with metabolic impairment as a result of effluent exposure), and food limitation (see Munkittrick et al., 2002).

2.2.2 Impact on Benthic Invertebrates

Benthic invertebrate communities are excellent indicators of the health of fish habitats, and variations in their structure are signs of habitat degradation due to contaminant or nutrient stressors. Pulp and paper mill effluents have been shown to affect the diversity and abundance of benthic invertebrate communities. The release of organic matter and other chemical compounds in effluents can significantly alter the receiving aquatic environment and lead to increases in nutrient enrichment, primary production, and toxic inhibitions, all of which can lead to detrimental impacts on benthic invertebrate communities.

The degradation of the large volumes of organic matter discharged by pulp and paper mills into the receiving environment can create anoxic conditions and smother the substrate, thereby impacting the invertebrate community (Pearson and Rosenberg, 1978; Bonsor et al., 1988). Anoxic conditions have been associated with lower invertebrate feeding rates, potentially lowering their growth and biomass; this poses a unique problem in northern rivers, where ice cover can restrict aeration (Lowell and Culp, 1999; Chambers et al., 2000a). Conversely, studies have also shown increases in growth, biomass, and reproductive output for invertebrates downstream of pulp and paper mills discharging substantial quantities of organic matter (Pearson and Rosenberg, 1978; Hall et al., 1991; Sibley et al., 1998). Increases in abundance, growth, and survival of benthic invertebrates have also been observed downstream of pulp and paper mills with effluents high in nutrients (Lowell et al., 1995, 1996, 2000; Dubé and Culp, 1996; Felder et al., 1998; Lowell and Culp, 1999, 2002; Culp et al., 2000a,b).

There have been a number of studies focusing on the effects of treated bleached kraft mill effluent on benthic invertebrate communities. Several studies have demonstrated increased invertebrate productivity at bleached kraft mills at low effluent concentrations (<2%) and explained these results, in part, by an increase in primary production (i.e., periphytic algae) (Hall et al., 1991; Lowell et al., 1995, 1996; Dubé and Culp, 1996; Culp et al., 2000b). At higher effluent concentrations (>5%), however, these same studies often measured either inhibitions or no changes in invertebrate growth. Lowell et al. (1995, 1996) and Dubé and Culp (1996) suggested that this inhibition might be due to other compounds in the effluent, including wood extractives. Both enrichment (e.g., nutrient) and inhibition effects (e.g., toxic compounds) can act simultaneously in an environment receiving effluent discharges; however, one effect could potentially mask the other (Lowell et al., 1995, 1996, 2000; Dubé and Culp, 1996; Culp et al., 2000a).

In a field mesocosm study by Lowell et al. (1995), mayflies exposed to 1% and 10% bio-treated bleached kraft mill effluent exhibited stimulatory growth under non-limiting food conditions. The authors suggested a number of mechanisms that might explain the increase of mayfly growth after exposure to effluent: (a) the presence of a hormonal or other growth stimulators, (b) an increase in the nutritive value of food, or (c) an increased feeding or palatability of the food. The phenomenon of low levels of toxic chemicals or other stressors leading to a stimulatory effect (i.e., increased growth) was termed *hormesis* in 1943 by Southam and Ehlrich (Lowell et al., 1995).

Bothwell and Stockner (1979) showed a reduction in effluent toxicity and nutrient effects at bleached kraft mills biologically treating their effluent with secondary treatment. The relationship between nutrient enrichment and toxic inhibitions is further illustrated in Figure 2-1.

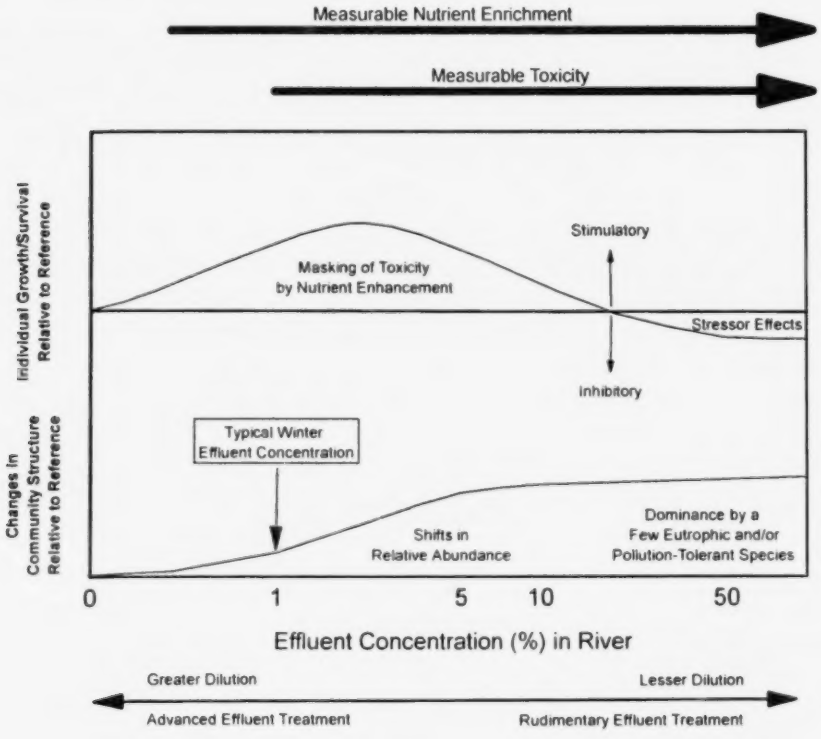


Figure 2-1: Schematic diagram illustrating the impacts of nutrients and toxicity, effluent treatment types, and receiving environment dilution on community structure and individual growth/survival rates of aquatic biota exposed to pulp mill effluent (obtained from Lowell et al., 2000)

Mill process may also have an impact on benthic invertebrate community structure and function, although the effect of process type is difficult to discern from the few published studies that are available. Thomas and Munteanu (1997) conducted a study in British Columbia in which effluent from a bleached kraft mill was shown to alter benthic invertebrate community structure, whereas exposure to effluent from a refiner mechanical mill did not lead to measurable changes in community structure, despite the addition of organic materials to the receiving reservoir. Lowell et al. (2000) and Lowell and Culp (2002) measured long-term (spanning 20 years) changes in community structure for benthic invertebrates exposed to secondarily treated bleached kraft mill effluent. On the other hand, in another study downstream of an unbleached

kraft mill with secondary treatment, the number of invertebrate taxa and species diversity were not impacted (Felder et al., 1998).

In a recent bleached kraft mill effluent study, Sibley et al. (2001) correlated benthic invertebrate community structure with distance from the effluent. The near shore (<300 m) had a predominance of pollution-resistant invertebrate species (i.e., tubificids), whereas no significant difference in community structure was observed at the far shore (>300 m) compared with the reference sites. As for Lowell et al. (2000), the final conclusion of this study was that benthic community structure can be impacted by pulp mill effluent and is a sensitive tool for monitoring environmental impacts.

Thus, more study is required to determine the relationship of mill process to the nature and magnitude of effects. The same holds true for evaluating the modifying effects of other key variables, such as effluent treatment and the habitat type of the receiving environment. Given the reality of the complex receiving water environments at most mills, evaluation of these modifying factors requires comparisons among large numbers of mills. To our knowledge, these kinds of comparisons have received very little study to date. The national assessment outlined in the following chapters provides a first step in furnishing this kind of information.

3. General Methods

3.1 Overall Data Preparation and Analysis

The national assessment of effluent-associated effects in the following chapters focused on three main components: the benthic invertebrate community field survey, the fish field survey, and sublethal toxicity testing done in the laboratory. This chapter outlines some of the general methods used to carry out the assessment, while the more specialized methods are further detailed in the following chapters, particularly for the sublethal toxicity assessment of effects on effluent quality.

A major task in assessing the field survey results was converting data into a standardized format that would allow comparisons among mills across the country, both within and between the invertebrate and fish population surveys. The main steps in data preparation and subsequent analyses are illustrated in simplified format in Figure 3-1. Most of the analyses were performed on data derived from electronic submissions of results from each mill. Extensive review of the submitted hardcopy reports was also necessary to obtain supplementary information required to make decisions on how to handle and interpret the electronic data.

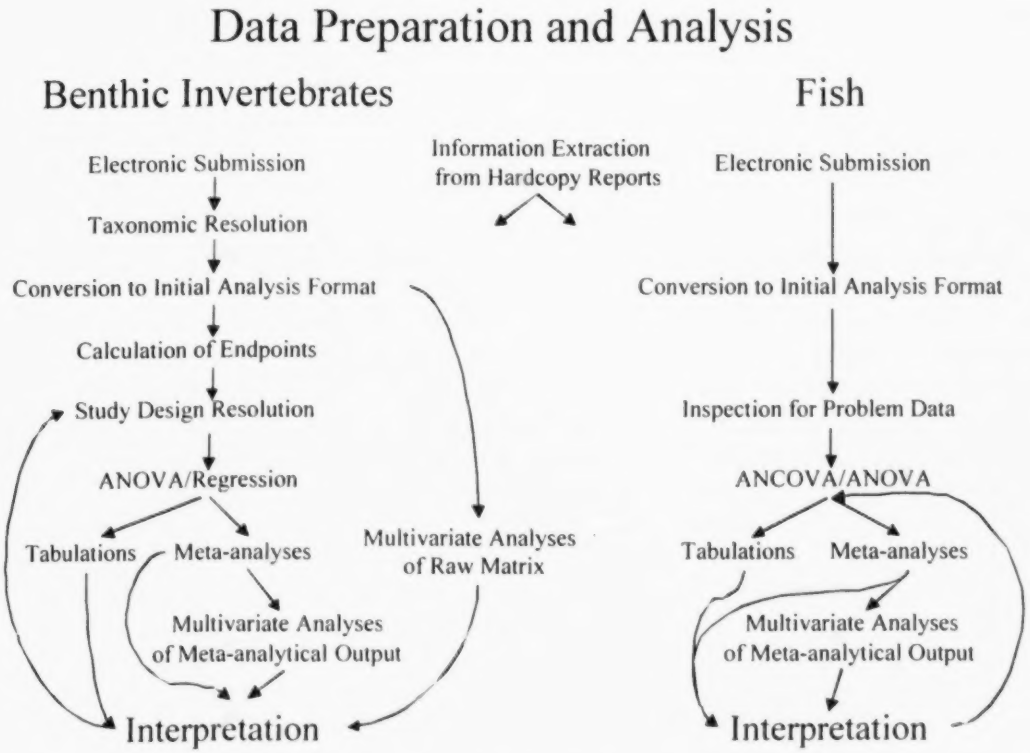


Figure 3-1: Simplified flow chart of main steps in data preparation and analysis

The primary steps common to both the invertebrate and fish analyses were as follows:

1) Electronic data were received and underwent initial checks for obvious errors (missing data fields, obvious data entry errors, etc.).

2) The data were stored in a standard database format and then exported into spreadsheets appropriate for analyses.

3) The spreadsheet data were analyzed via ANOVA, ANCOVA, or regression, depending on the study design, to determine the relationship between exposure to the effluent and the endpoint values for each mill (see following chapters). The ANOVA and ANCOVA analyses provided area means and standard deviations required for subsequent tabulations and meta-analyses of measured effects for the Control/Impact-type designs (some benthic invertebrate and all fish population survey studies). The regression analyses provided correlation coefficients for meta-analyses of the Gradient-type designs (remaining benthic invertebrate survey studies). These study designs and respective analyses are further detailed in Glozier et al. (2002) and Lowell et al. (2002).

4) The following chapters focus primarily on the final, synthetic stages of analysis described below: a) tabulation of exposure versus reference area comparisons across individual mills, b) meta-analyses, and c) multivariate analyses of the output of the meta-analyses. The meta-analyses were conducted using MetaWin (Rosenberg et al., 2000).

5) The final-stage synthetic analyses provided the basis for the overall interpretations of pulp mill effluent effects nationally.

6) Finally, the whole procedure was conducted iteratively, with initial analyses and interpretations providing feedback information required to improve data-handling and statistical procedures during subsequent rounds of analysis.

Although both required extensive manipulation, the invertebrate survey database and calculations were significantly more complex than those for the fish population survey, thus requiring additional major steps (Fig. 3-1). First, a wide variety of benthic habitats were sampled nationally, and, as a result, a very large number of taxa were recorded. This necessitated sorting through and recoding all the taxa into a standardized, web-based taxonomic format, the Integrated Taxonomic Information System. The Integrated Taxonomic Information System uses an international, digital taxonomic coding system co-ordinated, in Canada, by the Department of Agriculture and Agri-Food (see Godfray, 2002, for linkages to further information on this system). Second, whereas the fish population survey ANOVAs and ANCOVAs were basically run directly on the data submitted by the mills for individual fish, for the invertebrate survey it was first necessary to perform extensive data manipulations to calculate community-level endpoints, before subsequent ANOVA and regression analyses could be conducted (see Environment Canada, 2002, for calculation procedures). Third, the study designs used for the invertebrate survey were considerably more varied and complex than those used for the fish population survey (Glozier et al., 2002), requiring a large amount of further processing to facilitate standardization.

3.2 Synthetic Analyses

A national assessment of the findings of the EEM program requires the synthesis of the results of a large number of independent studies. A synthesis of this kind is generally undertaken using one or a combination of three basic approaches (Gurevitch and Hedges, 2001). The oldest and most commonly used approach is a narrative review, which qualitatively describes the findings of numerous studies by making generalizations based on the writer's evaluation of the available information. For example, this is the approach typically used by textbooks. Although this approach is more subjective, it does have advantages, including taking advantage of expert opinion to help interpret a number of lines of evidence that may not be conducive to a more quantitative synthesis. This report uses some elements of the narrative review, although the main emphasis is on the following two quantitative approaches. This was facilitated by the standardized approaches for data collection adhered to by the EEM program and by more recent developments in the field of quantitative syntheses.

The two main quantitative synthetic approaches used in the following chapters are 1) the tabulation of the results of individual mill studies (sometimes called the "vote-counting" approach) and 2) a more recent quantitative technique commonly called meta-analysis. These two basic approaches were also supplemented with multivariate analyses to look for overall patterns. Use of this combination of approaches proved to be valuable for conveying different aspects of the main findings of the national assessment.

Several of the figures in the following chapters tabulate the results of the individual mill studies by giving the percentage (or number) of studies finding significant changes for a particular endpoint when comparing exposure areas with reference areas (via ANOVA, ANCOVA, or regression). This is a method of summarization that is familiar to most readers and does not need much further explanation (see text accompanying the histogram figures in the following chapters). These figures were quite useful for illustrating several of the main findings. A limitation of tabulations of this kind is that the significance level of each study is dependent not only on the magnitude of effect, but also on sample size. For both the benthic invertebrate community and fish population surveys, considerable effort was made to ensure adequate replication to detect mill exposure areas exhibiting more pronounced effects (see further discussion in Section 4.5.3). Nonetheless, each mill had limitations on the sample size that was achievable. It has been shown that sample size limitations can reduce the ability of individual study tabulations to detect some effects that may be occurring (see Hedges and Olkin, 1985). Therefore, we also used the more recently developed meta-analytical techniques described below, which do not have the same limitations as individual study tabulations.

Practitioners of EEM studies at the individual mill level will note that the EEM program uses statistically significant differences between exposure and reference areas (termed an "effect") as a trigger (when confirmed in two cycles) for more focused monitoring to determine the cause, magnitude, and extent of potential impacts that a mill's effluent may be having on the receiving environment. This definition of effect is not meant to indicate that the significant difference was necessarily of a magnitude to be of ecological significance. At the level of this national assessment, though, the use of statistical significance as a trigger for further study for any

particular mill is not applicable. That is, this assessment is not meant to address the need for any particular mill to conduct more focused monitoring. Rather, the tabulations of significant differences for individual mills are used here simply to help illustrate national patterns. The issue of magnitude of effect and ecological significance is, however, further discussed in Section 4.5.3.

Meta-analyses statistically examine the magnitude of measured effects in a way that is independent of sample size and the scale of measurement used, facilitating the quantitative synthesis of a large number of independent studies (Rosenberg et al., 2000; Gurevitch and Hedges, 2001). In the national EEM assessment, these procedures provided the average magnitudes of effects for several mill categories (and for the country as a whole) and allowed tests of whether these magnitudes were greater than would be expected due to chance. That is, the analyses indicated, in a quantitative fashion, what general conclusions could be drawn regarding the nature of effluent-associated effects during Cycle 2. Furthermore, by essentially treating the separate studies as replicates, it was possible to look at questions that are very difficult to address at the level of the individual mill (e.g., the influence of habitat type, effluent treatment, and mill process on effluent effects in the field). Thus, the national EEM assessment addresses a number of issues that have been little (if at all) studied in the past.

For the purposes of analysis, the magnitude of effect in a meta-analysis is expressed as a standardized effect size. The main effect size measure in the following chapters is the Hedges' d, which is approximately equal to the difference between the exposure and reference area means divided by the pooled standard deviation (this value is multiplied by a correction factor that accounts for the effects of small sample sizes) (Rosenberg et al., 2000). That is, it is basically a measure of how large the measured effect is relative to background variability. For the benthic invertebrate community survey, it was also necessary to use a second kind of standardized effect size (the Fisher's z-transform) for some of the analyses, to enable the pooling of data from different study designs. Use of this second standardized endpoint is further discussed in Section 4.4.1. Positive values for either kind of effect size measure indicate a relative increase in the exposure area compared with the reference area, and negative values signal a relative decrease in the exposure area compared with the reference area. Thus, both the magnitude and direction of effect are important in interpreting the data.

In the following chapters, the main meta-analytical results are presented in the following summary format (Fig. 3-2). The standardized effect size is on the x-axis, with the vertical line representing a zero effect. The result for each mill grouping is presented as a horizontal 95% confidence interval about a vertical tick mark indicating the average effect size for that grouping of mills. Mill distributions to the right of the zero effect line indicate that the average effect associated with effluent exposure was an increase in the measured endpoint. Similarly, mill distributions to the left of the zero effect line indicate an effluent-associated decrease in the measured endpoint. The increase or decrease is statistically significant for the group as a whole if the 95% confidence interval does not overlap the zero effect line. Larger mill groupings (that are non-significant as a whole) can be composed of smaller subgroups, some or all of which may be significantly different from zero.

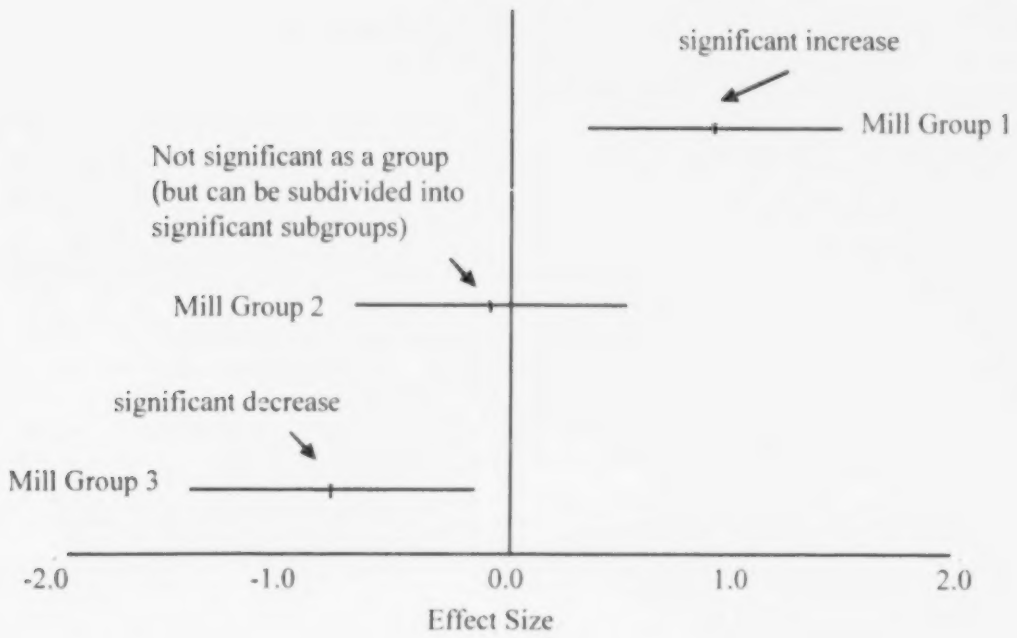


Figure 3-1: Meta-analysis summary figure

4. Benthic Invertebrate Community Survey

4.1 Introduction

The benthic invertebrate community survey comprises one of the two core field elements of the EEM program (the other being the fish population survey). During each cycle, the invertebrate survey was used to help determine the degree of habitat degradation due to organic enrichment or other forms of physical and chemical contamination and to provide an evaluation of the aquatic food resources available for fish.

Changes (in exposure areas relative to reference areas) in the following four core endpoints were used to assess effluent effects on benthic invertebrate community structure:

- total abundance (of individuals);
- taxon richness (number of taxa);
- Simpson's evenness (how evenly individuals are distributed among the taxa); and
- Bray-Curtis index of dissimilarity (change in overall community structure, particularly community composition, relative to reference communities).

The calculation and interpretation of these endpoints are further detailed in Environment Canada (2002).

The primary focus of this chapter is to provide a national evaluation and synthesis of the effects of pulp and paper mills on benthic invertebrate communities in Canadian receiving waters, as measured during Cycle 2 of the EEM program. Further, as outlined in Chapter 1, emphasis is placed on certain priority areas identified by the National EEM Science Committee, including the influence of habitat, mill process, and effluent treatment type on measured effects, as well as the evaluation of selected benthic program elements, as discussed in Section 4.5.

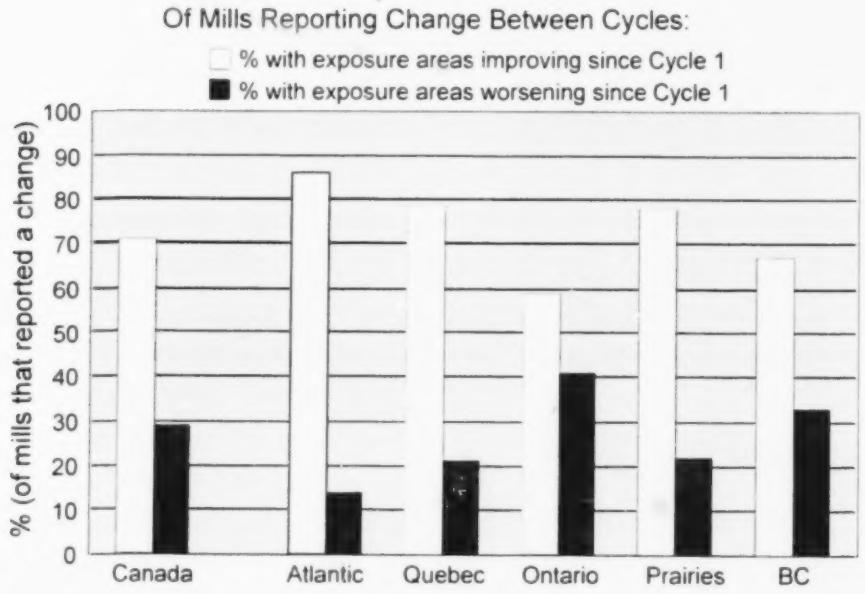
4.2 Comparison with Cycle 1 Results

In both Cycle 1 (Environment Canada, 1997d) and Cycle 2 (see Fig. 4-21 below), the majority of mills reported statistically significant differences in benthic invertebrate community structure between exposure and reference areas. Many of these differences may have been partly attributable to the confounding influence of non-mill-related factors, such as the presence of other effluents. A number of study design improvements were put into place following Cycle 1 (Environment Canada, 1997d), which helped to rule out many of the confounding influences, although some remained in Cycle 2.

In the interpretative reports submitted by each mill, some effort was made to compare Cycle 1 with Cycle 2 in terms of exposure versus reference area differences in invertebrate community structure. These changes in effluent-related effects between cycles were not consistently reported, in part due to changes in sampling procedures that would invalidate inter-cycle comparisons. Even so, approximately two-thirds of the studies reported notable changes in benthic conditions between cycles. These were based on both qualitative and quantitative

evaluations. Of the reports that noted changes, just fewer than three-quarters of the studies reported an improvement in benthic conditions between cycles (Fig. 4-1). This overall improvement was reported across the country, although there were some differences among regions. A smaller number of reports also indicated a worsening of benthic conditions, although some of these cases may have been due to changes in other non-mill-related factors. In many cases, the improvements in benthic conditions were attributed to improvements in mill effluent treatment practices, such as secondary effluent treatment, which led to significant decreases in effluent sublethal toxicity (see Chapter 7) and lessened eutrophication.

Figure 4-1: Percentage of interpretative reports reporting an improvement or worsening in benthic conditions in exposure areas relative to reference areas in Cycle 2 (of mills reporting a change)



4.3 Data Extraction and Preparation for Analysis

The benthic invertebrate survey data collected during Cycle 2 were integrated and synthesized nationally using the electronic data submitted by each mill. Some difficulties were encountered with the electronically submitted data, and these were dealt with as summarized below.

Following processing and editing, electronic data were available for 62 Control/Impact and 21 Gradient mill studies (see below).

Analysis and interpretation of these data on a national basis are ongoing, but the initial main findings are presented in the following sections. These first sets of analyses centred on the questions identified in Chapter 1 and were facilitated by focusing on certain key elements of the available data. To gain an initial understanding of overall effects, analyses were performed at the family level, as suggested by recent studies of the advantages of analyses at different taxonomic levels (Bowman and Bailey, 1997; Bailey et al., 2001). Where data were reported only at higher taxonomic levels, the lowest available taxonomic level was used. Future analyses of the Cycle 2 data will look at effects at lower taxonomic levels, which may yield further information on changes in community structure (Lenat and Resh, 2001).

Two principal kinds of benthic invertebrate study designs were used in Cycle 2: Control/Impact and Gradient, which are described in detail by Glozier et al. (2002) and Environment Canada (2002). In brief, the Control/Impact design entailed ANOVA comparisons of invertebrate community endpoint values between exposure and reference areas. The exposure areas were usually divided into near-field and far-field areas. This initial assessment focuses on near-field effects for the Control/Impact design in order to investigate the more pronounced effects that are occurring nationally. Future analyses will focus on effects farther from the mill outfalls. Analyses of Gradient design data incorporated regressions of endpoint values against distance from the outfall. The Gradient design results were also strongly influenced by near-field effects, with sampling stations closest to the outfall being equivalent to near-field areas and stations farthest from the outfall serving as surrogate reference areas (see discussion in Section 4.5.2). The Control/Impact ANOVA results and the Gradient regression results were incorporated into the meta-analyses and tabulations as outlined in Sections 4.4 and 4.5.

Following Cycle 1, a number of recommendations were made regarding how to improve study design, sampling methods, supporting information, and analysis for the benthic invertebrate community survey (Environment Canada, 1997d). Incorporation of these recommendations during Cycle 2 significantly improved the quality of the data collected during Cycle 2. Specifically, and probably most importantly, sampling station replication was increased to allow for more meaningful (and statistically powerful) comparisons of reference areas with exposure areas. Sampling effort was also usually concentrated into a single dominant habitat type at each mill, further increasing the resolving power of the study.

A greater variety of standardized study designs were used in Cycle 2 (e.g., different types of Control/Impact and Gradient designs) to allow for more site specificity. Each of the increased number of approved study designs was linked with an appropriate statistical procedure in the guidance document. This improved site-specific interpretations at each mill, although the increased number of study designs did make it more difficult to investigate national patterns in effects. Nevertheless, it was still possible to examine national patterns through the use of the synthetic analytical approaches described in Chapter 3 and in the following sections.

Other improvements in Cycle 2 included standardization of sampling methods, making the results of different studies more comparable. The measurement of supporting field measurements

was increased, although this improvement was offset by a widespread failure to report this information electronically. Data interpretation was also improved for some mills via a greater emphasis on graphical and tabular reporting of the recommended simple community descriptors, which helped with extraction of the main findings from each study. It is highly recommended that this method of data presentation is used more widely in future interpretative reports.

Although these improvements in data collection and analysis greatly enhanced the quality of the Cycle 2 data, some difficulties were encountered during the electronic data submission process. A particularly difficult problem was the entering of incorrect, outdated, and inappropriate taxa. As discussed in Chapter 3, this was resolved by recoding all the data into an international, standardized, web-based format, the Integrated Taxonomic Information System (see Godfray, 2002, for linkages to further information on this system).

Other problems encountered with the electronically submitted data included the following:

- incorrect station identifications, making it difficult to tell where the data were collected;
- labeling reference stations as exposure stations, and vice versa;
- the number of stations in electronic files not matching the number of stations in printed interpretative reports;
- raw numbers in electronic files not matching numbers in printed interpretative reports;
- insufficient identification of which reference and exposure stations were in comparable habitats, so that it was not possible to do a reference versus exposure comparison;
- not distinguishing data collected from natural substrata versus artificial substrata;
- inclusion of (and not identifying) historical station data that could not be compared with Cycle 2 station data;
- inclusion of (and not identifying) stations that were not included in the original interpretative report analyses because of problems that had been discovered with those stations; and
- failure to enter the requested supporting environmental variables, so that it was not possible to look at their effects on benthic responses.

The preceding electronic data submission problems were, however, mostly the exception rather than the rule. Given the massive number of data that were submitted, most of the data were of good quality, thus permitting us to conduct the national synthesis described in the following sections. We manually corrected the above errors when they were discovered. In cases where correction was not possible, the data were excluded from analysis. We expect these problems to be greatly reduced in future cycles via improvements in study procedures and by the provision of standardized menus for use during electronic data entry.

We also compared the data submitted in the interpretative reports with the data submitted electronically, to evaluate consistency. Due to the complexity of the benthic invertebrate survey designs, a wide variety of methods of analysis and presentation were used among the different report submissions. Thus, it was difficult to summarize the reports in a consistent fashion. The majority of reports did, however, provide some form of statistical comparison of exposure with reference areas for two standard endpoints — total abundance and taxon richness. Therefore, we

compared those mill consultant calculations with equivalent calculations using the electronically submitted data.

Figure 4-2 gives the percentage of mills where statistically significant effects were calculated for one, both, or neither of the two endpoints. This figure shows that the interpretative report and electronic data agreed quite well. The slight differences that were found were largely a result of the data submission problems described above. The greater incidence, relative to the fish population survey results (Fig. 5-11 in Chapter 5), of mills reporting no significant differences is likely largely due to the fact that only a single set of benthic endpoints was calculated for each mill, whereas up to four sets of fish endpoints were calculated for each mill (up to two species and two sexes). In addition, it was possible to include three kinds of endpoints in the fish electronic versus report comparison (condition, liver, gonad), compared with two kinds of endpoints for the benthic comparison (abundance, taxon richness). That is, for Figure 5-11 in Chapter 5, there could be up to 12 comparisons per mill (2 species × 2 sexes × 3 endpoints = 12), compared with only two comparisons per mill in Figure 4-2 below (two endpoints). The additional number of comparisons per mill in Figure 5-11 of Chapter 5 decreased the probability that none of them would be significant.

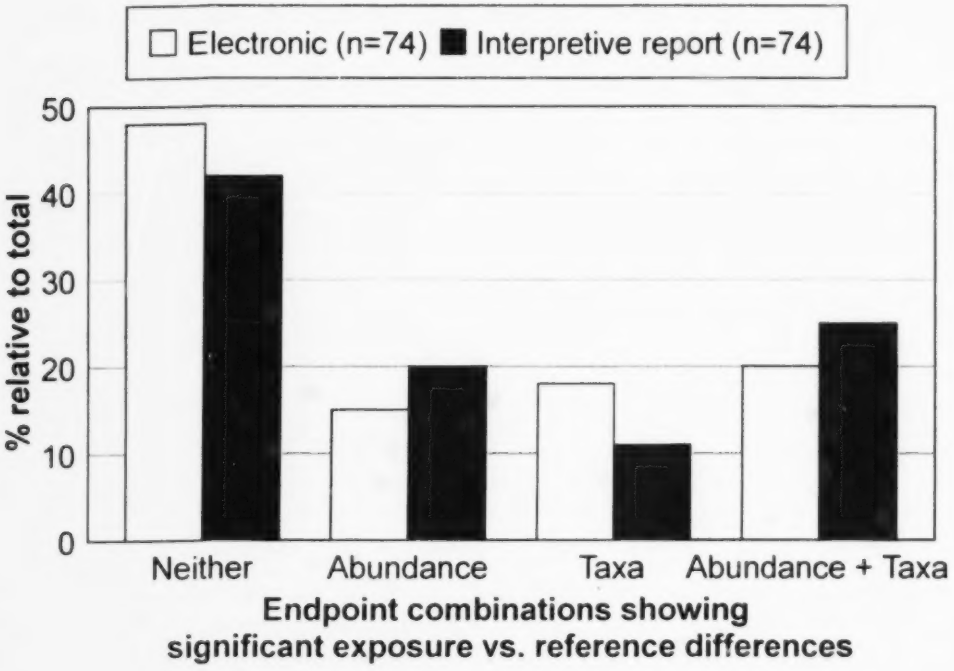


Figure 4-1: Comparison of electronic versus interpretative report data: Percentage of mills where statistically significant effects were measured for one, both, or neither of the two endpoints (abundance, taxon richness)

4.4 *National Trends*

4.4.1 *Overall Response Patterns*

Previous studies have provided insight into the expected effects of pulp mill effluent on benthic invertebrate communities (Lowell et al., 1995, 1996, 2000; Lowell and Culp, 1999, 2002; Culp et al., 2000a,b), and the overall nature of these interacting effects is summarized in Figure 4-3 and the following discussion. This figure illustrates expected changes in abundance, taxon richness, and community structure with changing effluent quality. In this context, "effluent quality" refers to that experienced by organisms exposed in the field, as distinguished from effluent effects measured under controlled conditions in the laboratory. The x-axis progresses from better quality effluent and less deleterious effects on the left to poorer quality effluent and more deleterious effects on the right. More deleterious effects may also be associated with past historical effects as
well (such as the smothering effects of fibre mats generated over many years of mill operation).

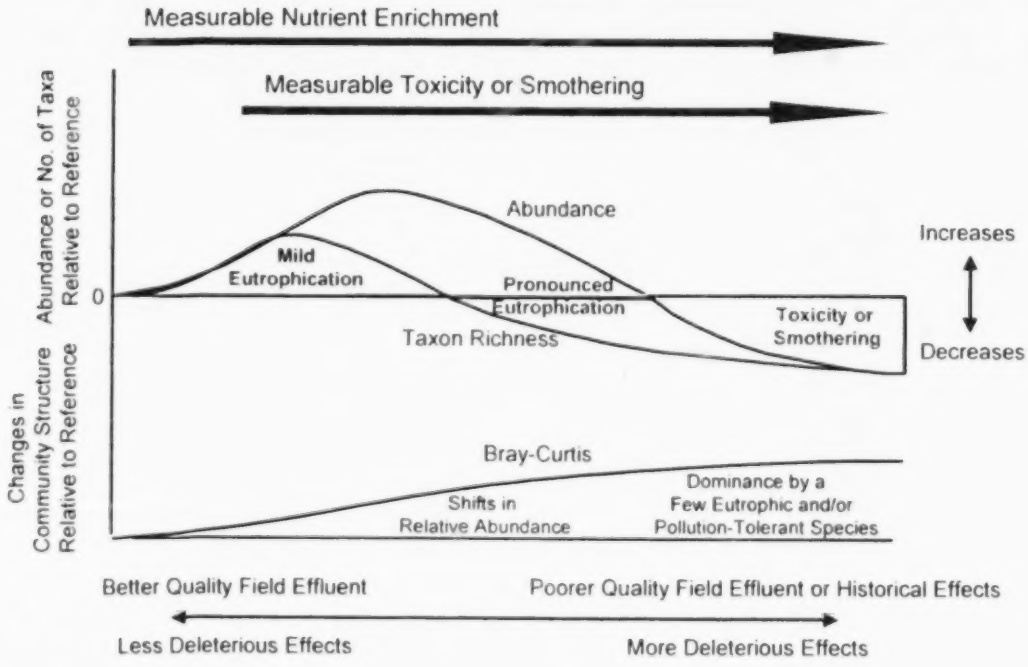


Figure 4-1: Predicted response patterns (see text; modified from Lowell et al., 2000; see also Culp et al., 2000a)

In general, nutrient enrichment (or eutrophication) increases from left to right, along with increasing toxicity or smothering effects (Fig. 4-3). Nutrient enrichment can often be measured at lower effluent concentrations than toxicity, and toxic effects are often masked by eutrophication at low to medium concentrations. Mild eutrophication is typified by increases in both total abundance and taxon richness. Progressing to the right, moderate eutrophication is typically associated with lessened increases in taxon richness, although further increases in abundance may still occur. More pronounced eutrophication is commonly associated with decreases in taxon richness, even while abundance is still greater than that found in reference areas. Finally, decreases in both taxon richness and abundance are typically a sign of overall inhibitory effects, such as toxicity or smothering.

In the EEM program, changes in invertebrate community composition are measured by changes in the Bray-Curtis index of dissimilarity. The value of this index increases as poorer quality effluent leads to greater changes in community composition. This is illustrated at the bottom of Figure 4-3, although it should be emphasized that changes in community composition are not always tied to changes in total abundance and taxon richness. This is because there are many different ways in which benthic communities may be affected in the field. Due to complex direct and indirect effects, effluent exposure may lead to pronounced effects on community composition without large effects on abundance or taxon richness, and vice versa.

Inspection of the grand means from the meta-analyses reveals the average overall effect on benthic invertebrate communities associated with exposure to effluent (Figs. 4-4 and 4-5). These means represent the averages across the country for all habitat types. Also shown are the 95% confidence intervals for these averages, indicating the amount of variability among mill findings. Figure 4-4 shows the grand means for mills using a Control/Impact study design. The Control/Impact results are shown separately, both because this was the most commonly used design and because this design appeared to have higher resolution than the Gradient design for identifying effects. This latter point is further discussed in Section 4.5.2, which also discusses the advantages of the Gradient design.

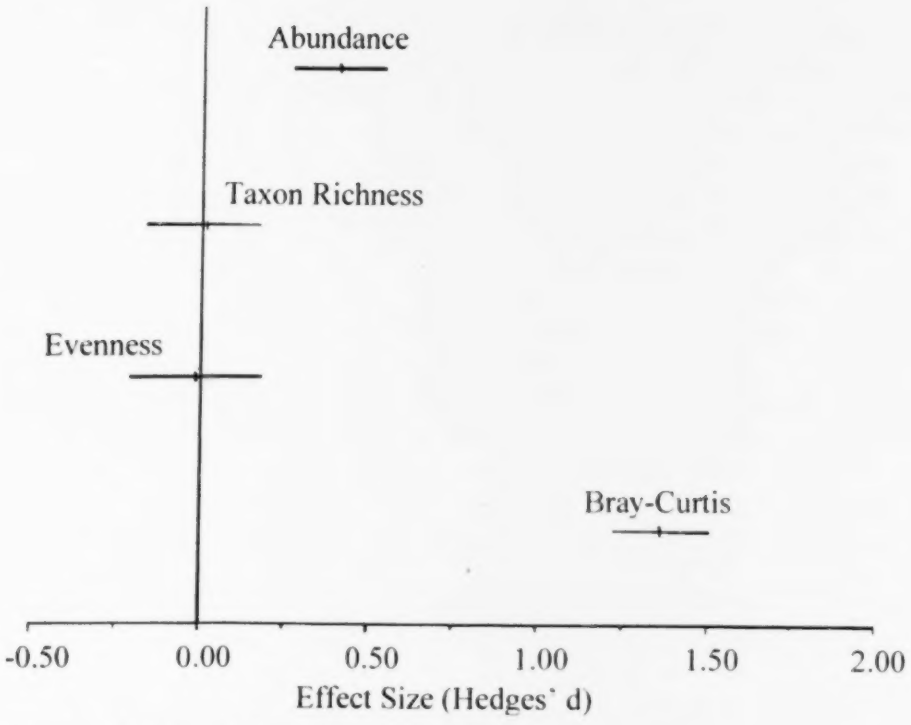


Figure 4-2: Control/Impact mill grand means

The total abundance grand mean was highly significantly positive (95% confidence interval was well above the zero effect vertical line) (Fig. 4-4). This means that the most common effect associated with exposure to effluent was an increase in invertebrate numbers. This is a well-known eutrophication-type response to exposure to pulp mill effluent, due to the nutrient enrichment effects of the phosphorus, nitrogen, and organic content of the effluent, a phenomenon that has been documented in North America (Hall et al., 1991; Dubé and Culp, 1996; Lowell et al., 2000; Chambers et al., 2000b; Culp et al., 2000a,b; Lowell and Culp, 2002) and Europe (Hansson, 1987; Feder and Pearson, 1988).

In contrast to abundance, the taxon richness grand mean was close to zero (Fig. 4-4). This reflects a more even distribution of mills with increases, mills with decreases, and mills with no significant change in number of taxa. The abundance and taxon richness distributions, taken together, indicate that the most common effect across the country was an increase in abundance and either an increase or no change in taxon richness, which is indicative of mild eutrophication (as illustrated in Fig. 4-3). The smaller number of mills with increases in abundance and decreases in taxon richness were showing typical signs of more pronounced eutrophication.

The evenness grand mean was also close to zero, also reflecting a more even distribution of mills with increases, decreases, or no significant change in evenness (Fig. 4-4). On the other hand, the Bray-Curtis grand mean was highly significantly positive. This shows that there were clear changes in benthic invertebrate community composition in exposure areas across the country. Because of the way in which the index is calculated, any changes in community composition in

exposure areas result in positive (not negative) Bray-Curtis values, with larger effects yielding larger Bray-Curtis values. The particularly high level of significance for the Bray-Curtis endpoint was, in part, due to the high sensitivity of this endpoint. This high sensitivity results because this endpoint is sensitive not only to changes in total abundance and taxon richness, but also to changes in community composition that the other three endpoints do not detect. This makes it more sensitive to changes in overall community structure than the other three endpoints (see Section 4.5.1 for further discussion). We are currently evaluating a series of new statistical approaches that combine multivariate analyses of the raw benthic matrices with meta-analysis to identify which taxa are most responsible for the changes in community structure measured by the Bray-Curtis endpoint.

It should be further noted that the majority of mill effects were well below a standardized effect size of two for most of these endpoints. This is of significance for the Hedges' d measure of effect size used for the Control/Impact mills (Fig. 4-4 and following Control/Impact figures). As described above, this measure is approximately equal to the difference between the exposure and reference means divided by the pooled standard deviation. That is, Hedges' d approximately illustrates how far apart the exposure and reference means are in terms of standard deviations. Put another way, Hedges' d expresses the magnitude of effect relative to background variability. Since most of the Cycle 2 benthic invertebrate studies were designed to have only sufficient power to detect an effect size as large as two standard deviations or more, Figure 4-4 illustrates that, on average, only mills with especially large effects would have measured a statistically significant effect for a given endpoint when performing ANOVA comparisons on an individual mill basis. The one exception is the Bray-Curtis endpoint for certain groupings of Control/Impact mills, as seen later in this section. This is due to the high sensitivity of the Bray-Curtis endpoint (see Section 4.5.1). Section 4.5.3 discusses some further issues related to these latter points.

It was also possible to combine the results from the Gradient design mills with those from the Control/Impact design mills (Fig. 4-5). Whereas the magnitudes of effects at Control/Impact mills were measured as proportional differences between reference and exposure areas, this was not possible for Gradient mills, because the latter did not sample true reference areas. Rather, the magnitudes of effects for Gradient mills were determined by regressing endpoints against distance from the mill outfall and calculating the correlation coefficient. Thus, it was necessary to convert the two kinds of results (two study designs) into a common currency. The results for both study designs were combined by converting either proportional reference/exposure differences (Control/Impact mills) or correlation coefficients (Gradient mills) into a single standardized effect size, termed a Fisher's z-transform. Although this latter effect size measure does not have the same intuitive appeal as the Hedges' d (which basically expresses effect size in terms of standard deviation units), it does allow the results from the two study designs to be combined.

The overall effects when combining all the mills (both Control/Impact and Gradient; Fig. 4-5) were similar to those for Control/Impact mills alone (Fig. 4-4). The abundance grand mean was highly significantly positive, while the taxon richness and evenness grand means were close to zero. The Bray-Curtis endpoint was not calculated for the Gradient design mills because its calculation requires a comparison with a reference area, and reference areas were not sampled as part of the Gradient design. It would be possible to calculate an approximate Bray-Curtis

endpoint for Gradient mills by treating sampling stations farthest from the mill outfall as approximately equivalent to reference. Although this latter method of calculation requires some additional assumptions, it does have some value and will be included in future analyses of EEM data.

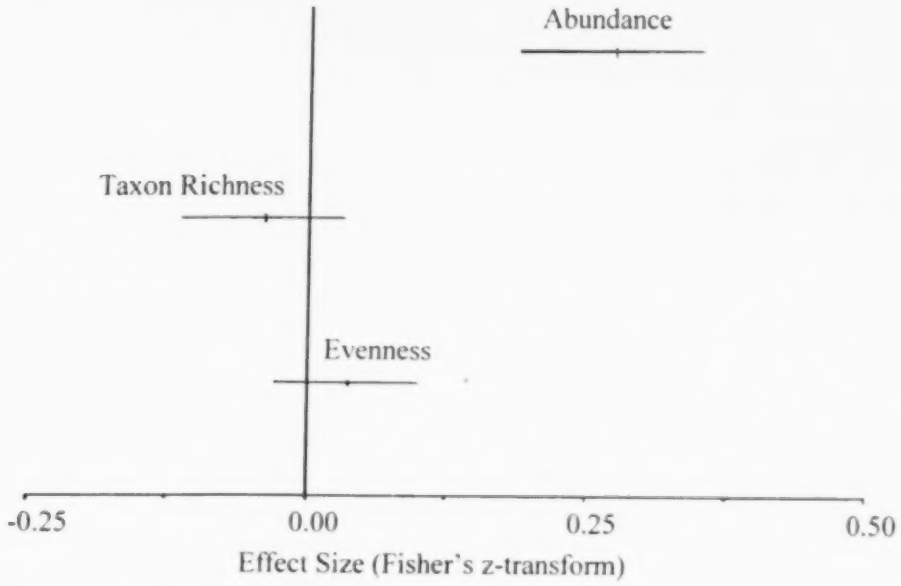


Figure 4-3: Control/Impact + Gradient mill grand means

Thus, the addition of Gradient mills to the analyses did not change the interpretation of overall effluent effects as averaged across the country — that is, the overall predominance of mild to moderate eutrophication effects. The taxon richness mean was shifted slightly to the left and the evenness mean was shifted a similar amount to the right, and this likely reflects the influence of marine/estuarine mills (see Section 4.4.2), which made much greater use of the Gradient design.

Whereas the meta-analyses were performed on standardized measured effect sizes (Hedges' d and Fisher's z-transform) to allow statistical comparisons among the mills, it is also useful to look at the distribution of the non-standardized measured effect sizes for the Control/Impact mills (Figs. 4-6 to 4-9; Tables 4-1 to 4-4). This helps to provide a further appreciation for the distribution and range in magnitude of the measured effects. The common way of calculating these latter effect sizes is as the exposure area mean minus the reference area mean, expressed as a percentage of the reference area mean.

The abundance endpoint showed the most extreme range in exposure-associated effects, extending from decreases of almost 100% to increases of over 2000% (Fig. 4-6). The majority of the measured effects for abundance ranged from decreases of 60% to increases of 900%. Measured effects for taxon richness ranged from decreases of 70% to increases of 150%, with most falling between -40% and +40% (Fig. 4-7). Evenness showed a distribution similar to (but slightly broader than) taxon richness, with the extremes ranging from decreases of just over 60%

to increases of 190% (Fig. 4-8). Due to the way in which it was calculated, the effect sizes for the Bray-Curtis measure of changes in community composition were mostly positive (Fig. 4.9). They ranged from approximately zero (no change in community composition) to increases of up to 500%, with most falling within the 0–250% range. For further comparison, Tables 4-1 to 4-4 list the greatest statistically significant increases and decreases in abundance and taxon richness that were measured in exposure areas during Cycle 2. As these figures and tables show, the magnitudes of effects associated with effluent exposure were often quite large, although these magnitudes should be considered within the context of the range of variability typically seen in reference areas. See Section 4.5.3 for a further discussion of the interpretation and application of these measured effects.

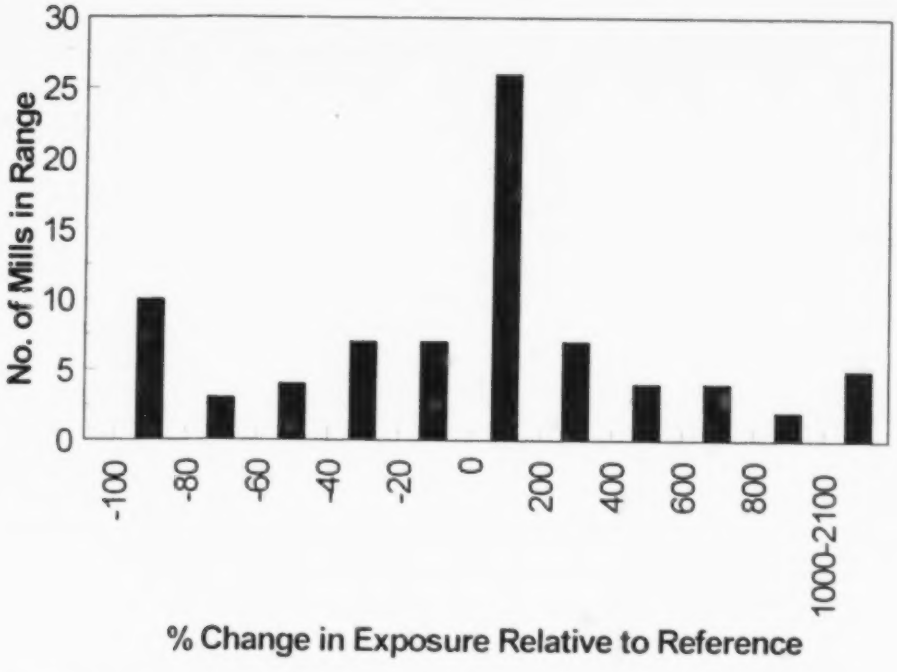


Figure 4-4: Abundance measured effect size distribution

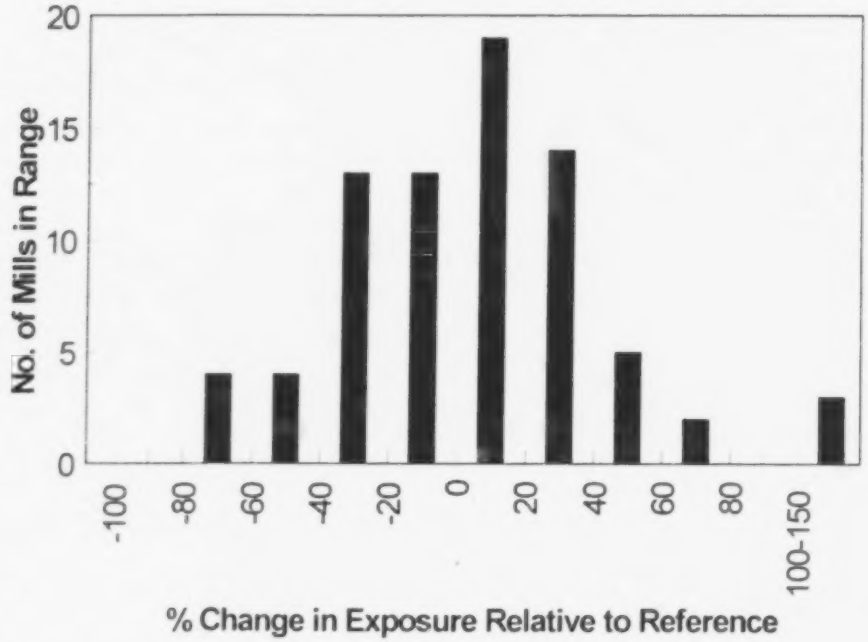


Figure 4-5: Taxon richness measured effect size distribution

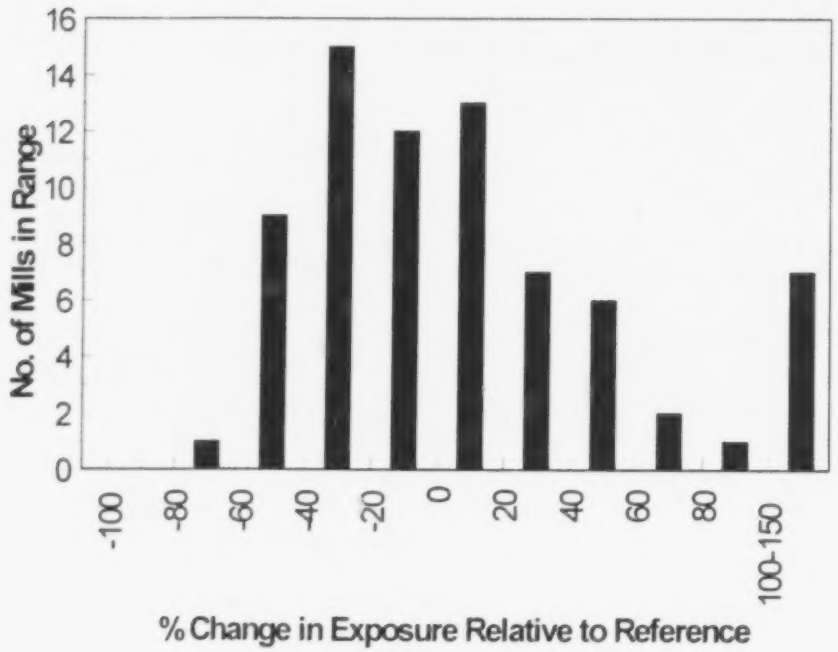


Figure 4-6: Evenness measured effect size distribution

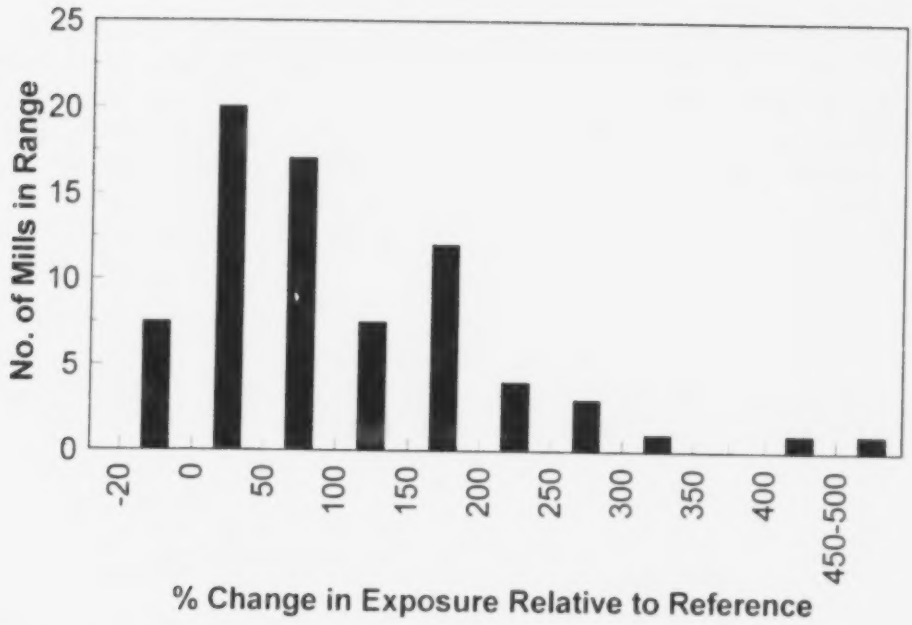


Figure 4-7: Bray-Curtis measured effect size distribution

Table 4-1: Greatest significant increases in abundance

Mill Rank	Abundance Effect Size	No. of Taxa Effect Size	Bray-Curtis Effect Size	Evenness Effect Size
1	2020.2	-35.7 (NS)	255.9	-15.4 (NS)
2	1947.6	150.0	493.8	-33.5 (NS)
3	1344.3	51.5	196.3	-24.9 (NS)
4	828.7	52.8	183.1	-45.1 (NS)
5	708.7	8.1 (NS)	47.1	-27.5 (NS)
6	669.8	-25.9	216.6	-44.1
7	659.7	-52.4	154.9	10.1 (NS)
8	432.1	105.1	120.8	-54.6
9	410.5	53.8	404.0	-29.9
10	374.8	29.3 (NS)	99.3	-19.3 (NS)

NS = not significant

Table 4-2: Greatest significant decreases in abundance

Mill Rank	Abundance Effect Size	No. of Taxa Effect Size	Bray-Curtis Effect Size	Evenness Effect Size
1	-94.2	-71.8	242.9	175.1
2	-89.5	-51.7	316.6	191.3
3	-80.4	-65.9	209.7	100.7

Table 4-3: Greatest significant increases in taxon richness

Mill Rank	No. of Taxa Effect Size	Abundance Effect Size	Bray-Curtis Effect Size	Evenness Effect Size
1	150.0	1947.6	493.8	-33.5 (NS)
2	105.1	432.1	120.8	-54.6
3	62.4	271.0	57.9	-9.9 (NS)
4	53.8	410.5	404.0	-29.9
5	52.8	828.7	183.1	-45.1 (NS)
6	51.5	1344.3	196.3	-24.9 (NS)
7	28.1	-89.6 (NS)	179.1	55.9
8	23.9	17.7 (NS)	290.3	-23.3
9	20.6	254.8	92.2	-32.7
10	18.3	76.4	46.9	7.4 (NS)

NS = not significant

Table 4-4: Greatest significant decreases in taxon richness

Mill Rank	No. of Taxa Effect Size	Abundance Effect Size	Bray-Curtis Effect Size	Evenness Effect Size
1	-72.7	-96.3 (NS)	253.7	175.4
2	-71.8	-94.2	242.9	175.1
3	-65.9	-80.4	209.7	100.7
4	-52.4	659.7	154.9	10.1 (NS)
5	-51.7	-89.5	316.6	191.3
6	-41.0	1176.3 (NS)	124.9	-30.3 (NS)
7	-37.8	-78.6 (NS)	13.2 (NS)	22.3 (NS)
8	-25.9	669.8	216.6	-44.1 (NS)
9	-24.8	-25.0 (NS)	98.4	43.9 (NS)
10	-22.8	9.1 (NS)	138.8	-18.1 (NS)

NS = not significant

4.4.2 Habitat

The magnitude and direction of effluent-associated effects on benthic invertebrate communities varied significantly among different habitat types. The numbers of mills in each habitat category for the benthic invertebrate data set are given in Table 4-5. To provide sufficient sample sizes (i.e., number of mills in each category), marine and estuarine mills were pooled for the Control/Impact meta-analyses.

Some of the most notable effects were apparent when considering changes in total abundance and taxon richness together (Fig. 4-10). Abundance was significantly increased in both erosional and depositional riverine habitats. The distribution of mill responses for taxon richness mostly ranged from increases to no significant change. This combination of responses (see Fig. 4-3) is

indicative of an average effect of mild eutrophication in river exposure areas. In contrast, both abundance and taxon richness decreased in marine/estuarine habitats (Fig. 4-10), a response that is indicative of an average toxicity or smothering effect. In fact, the magnitude of the decrease in the taxon richness mean for marine/estuarine habitats was close to two standard deviations. The distribution for lacustrine habitats was intermediate between that for riverine and marine/estuarine habitats. It should be noted that these are national averages and that not every mill fell into the average response patterns identified for each of the above habitat types.

Table 4-1: Number of mills in different habitat categories

Habitat Type	Number of Control/Impact Mills	Number of Control/Impact + Gradient Mills
Riverine — erosional	33	34
Riverine — depositional	19	21
Lacustrine	4	7
Marine/estuarine	6	–
Marine	–	13
Estuarine	–	8

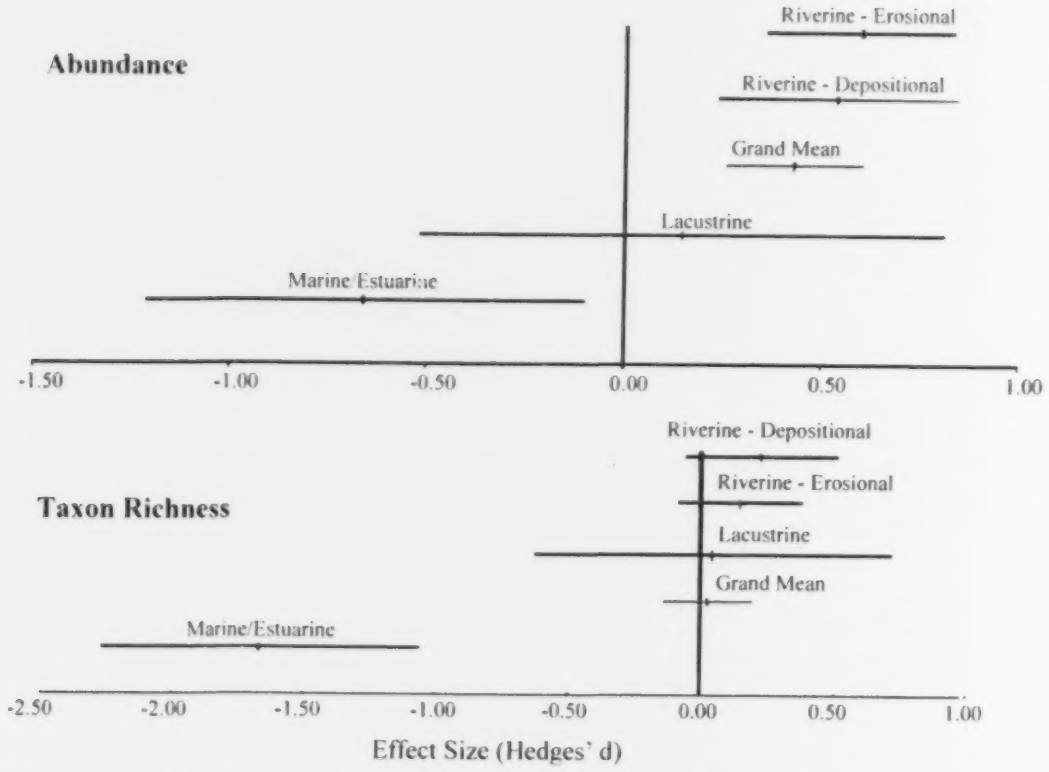


Figure 4-1: Abundance and taxon richness by habitat (Control/Impact mills)

The addition of Gradient mills to the analyses (Fig. 4-11) did not greatly change the overall pattern seen above for Control/Impact mills. Rivers still showed eutrophication effects, and lakes maintained their intermediary position. Inclusion of the Gradient mills did, however, provide enough mill replication to distinguish effects seen in marine versus estuarine habitats. This additional information revealed that abundance effects in estuarine habitats were intermediate between those in freshwater and marine habitats, in keeping with the intermediate physical characteristics of estuarine habitats.

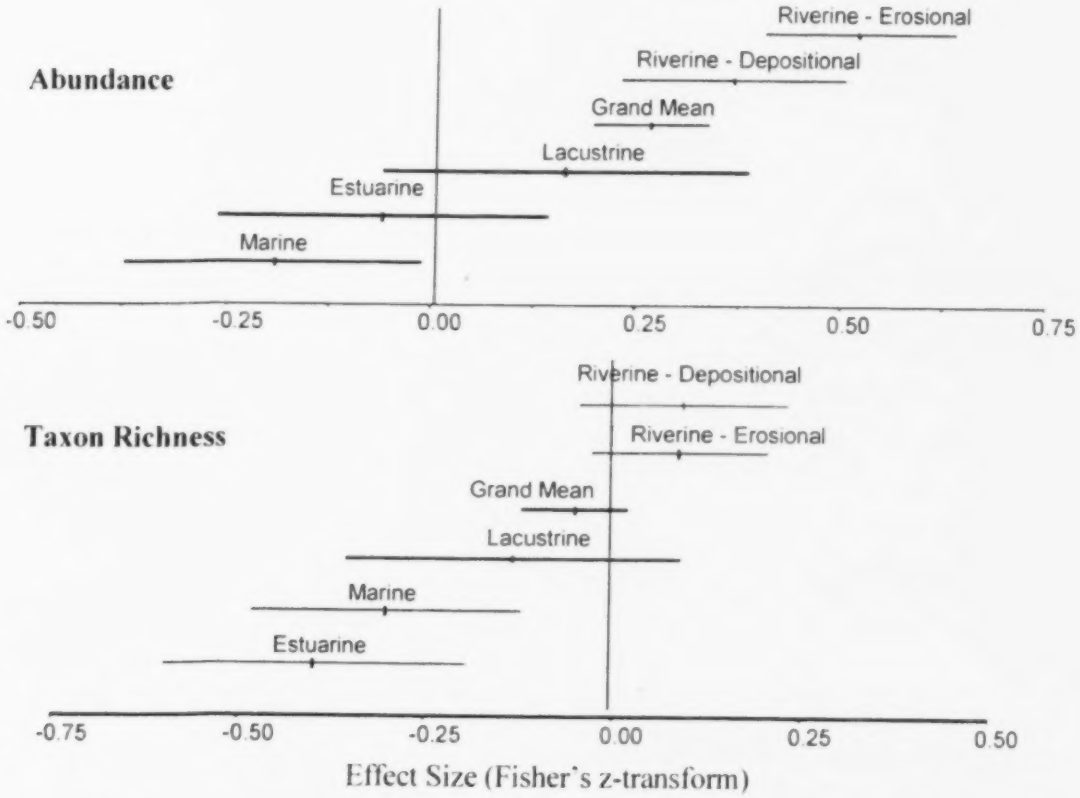


Figure 4-2: Abundance and taxon richness by habitat (Control/Impact + Gradient mills)

As discussed above, mild to moderate eutrophication effects have been well documented for mills discharging into riverine habitats (Culp et al., 2000a,b; Lowell et al., 2000; Lowell and Culp, 2002). The inhibitory response seen in estuarine and, particularly, marine habitats is likely due to a combination of factors. In the Atlantic region, the inhibitory effect may be partly because the sampling stations are located closer to the outfall for marine mills, although the inconsistent electronic reporting of distance to the outfall makes it difficult to fully evaluate this partial explanation. In the Pacific region, a similar inhibitory effect was seen for marine mills, even when sampling stations were located quite far from the outfall. Furthermore, nutrient enrichment effects were seen at freshwater mills even when sampling stations were located quite close to the outfall.

A likely cause of the inhibitory effects observed at several of the marine mills is the habitat deterioration due to historical deposition of fibre mats, which can sometimes be quite thick. If this was the predominant cause of the overall inhibitory effect seen for marine mills, this would suggest that historical fibre mat deposition is a greater problem for marine than for riverine habitat mills. In keeping with this hypothesis, a reduced (relative to marine) historical fibre accumulation in rivers may be due to the regular flushing of riverine benthic habitats by the freshet that occurs every spring/summer in a large number of Canadian rivers. Further work is required to determine whether other factors (such as differences in taxa sensitivities to effluent exposure) also contributed to the different responses in marine habitats.

The magnitude and direction of effects on evenness also varied among the different habitat types (Fig. 4-12). For freshwater depositional habitats (lakes and depositional rivers), the 95% confidence intervals overlapped zero to a fairly large degree. In erosional riverine habitats, however, evenness was nearly significantly reduced overall in exposure areas, which may reflect the sensitivity of organisms in erosional habitats to even mild eutrophication.

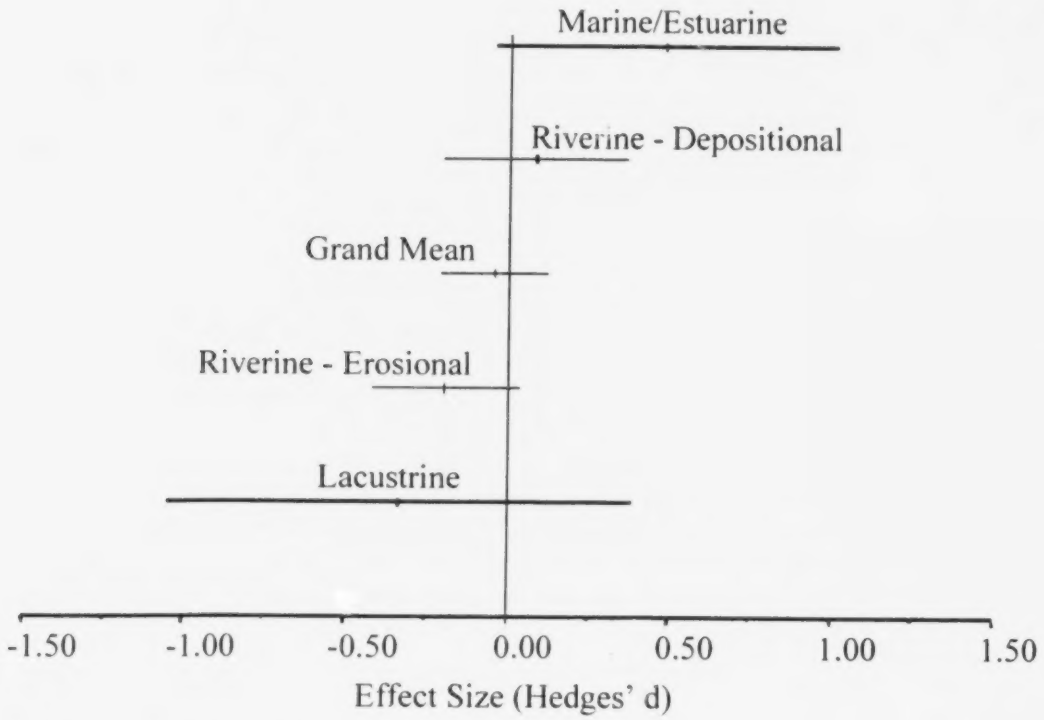


Figure 4-3: Evenness by habitat (Control/Impact mills)

Interestingly, the addition of the Gradient mills revealed that evenness was significantly increased in estuarine exposure areas (Fig. 4-13). This kind of stress effect is less commonly cited than decreases in evenness. The data suggest that toxicity/smothering effects may have led to the loss of more rare taxa, leaving behind a fewer number of more evenly distributed, pollution-tolerant taxa.

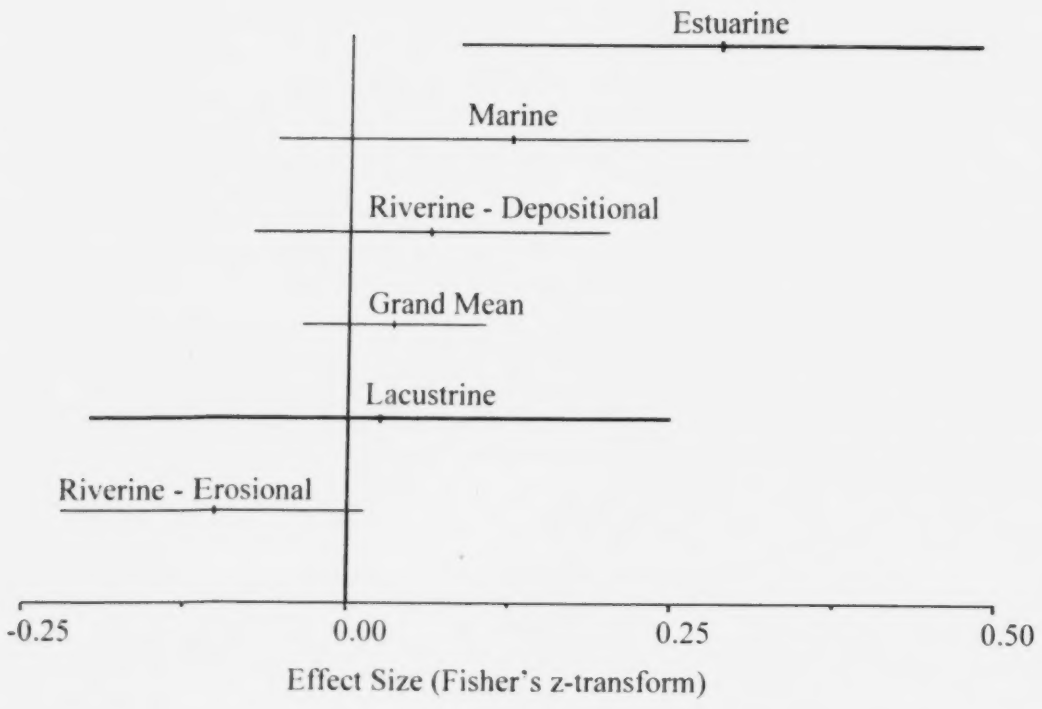


Figure 4-4: Evenness by habitat (Control/Impact + Gradient mills)

Of the four endpoints, the largest exposure area effects for the habitat subgroups were observed for the Bray-Curtis endpoint (Fig. 4-14). As discussed in Section 4.5.1, this gauge of changes in community composition was the most sensitive of the measured endpoints. Of the different habitat types, the largest exposure-associated changes in community composition were measured in lacustrine habitats (over three standard deviations), although any conclusions are tentative, given the small sample size. This finding is noteworthy, in that measured lake effects were more intermediate for the other endpoints. This suggests that, under different environmental conditions, effluent-associated effects may be manifested in diverse ways. The different patterns detected by the four endpoints underscore the value of measuring changes in community structure in different ways (i.e., using multiple endpoints) to reduce the probability of missing important (and sometimes quite large) effects.

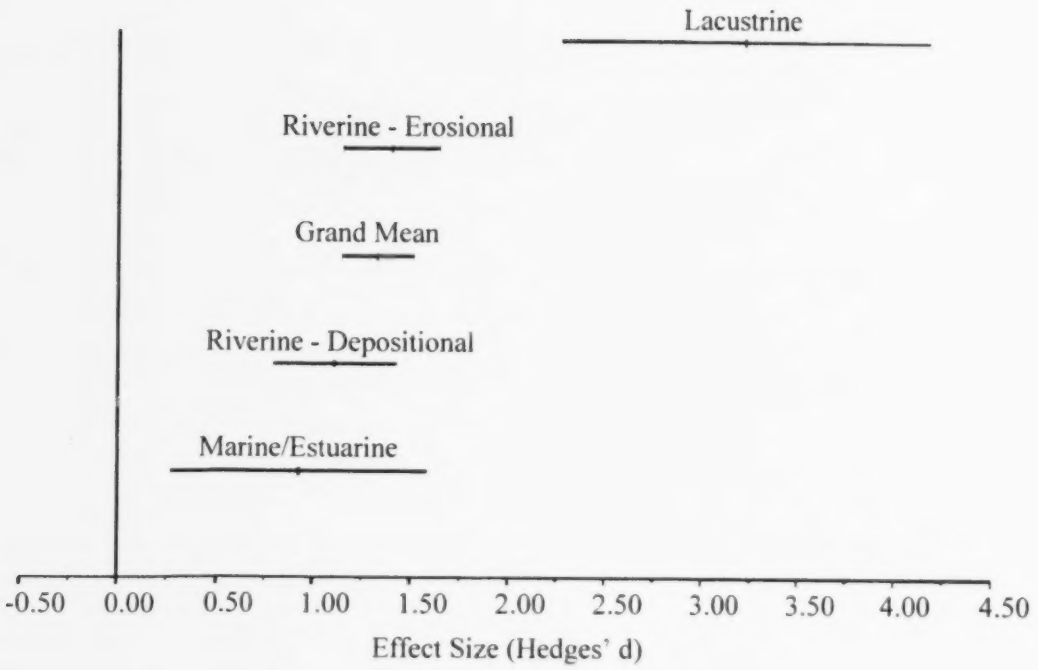


Figure 4-5: Bray-Curtis by habitat (Control/Impact mills)

4.4.3 Effluent Treatment and Mill Process Types

Pulp mills in Canada use a wide variety of effluent treatment and mill process types. Considered in great enough detail, each individual mill is unique, especially in terms of mill process. Nevertheless, the mills can be grouped into certain broad categories, as given in Table 4-6. It can be quite valuable to determine whether certain effluent treatment and mill process types are associated with greater or lesser effects in the environment or whether the different groupings are all basically equivalent in terms of their environmental effects. We investigated possible patterns in these effects by conducting meta-analytical comparisons among the different effluent treatment and mill process types. The numbers of mills in each group for the benthic invertebrate data set are also given in Table 4-6. There were insufficient numbers of mills in many of the groupings to allow firm conclusions to be drawn at present for the smaller groupings. The data did suggest some large environmental effects associated with some of the smaller groupings, however, which may be deserving of further study.

Table 4-1: Number of mills in different effluent treatment and mill process categories

	Number of Control/Impact Mills	Number of Control/Impact + Gradient Mills
Effluent Treatment Type		
Activated sludge treatment	35	45
Aerated stabilization basin	18	25
Sequential batch reactor	4	5
Primary	3	4
Other	2	4
Mill Process Type		
Kraft	19	26
Thermomechanical	15	20
Chemi-thermomechanical	6	8
Mechanical/sulphite	5	6
Mechanical (other)	6	6
Semi-chemical	2	3
Recycled	4	6
Paper mill only	5	5
Other	0	3

In the meantime, we present here the findings for the groupings represented by large numbers of mills. For comparative purposes, the grand means across all the treatment and process types are also given in each of the figures in this section. The two most common effluent treatment types were activated sludge treatment and aerated stabilization basin. The two most common mill process types were kraft and thermomechanical. Because these most common types represented the majority of the mills, the patterns in the following figures mostly reflect the overall countrywide pattern. Exposure area effects for the two most common effluent treatment types were fairly similar to each other. Exposure area effects for the two most common mill process types were more different from one another. Although it is possible that these latter differences were due to co-occurring variables other than mill process type, the large sample sizes (i.e., large numbers of mills in both subgroups) suggest that any such co-occurring variables would need to be fairly tightly linked to mill process type.

Exposure area abundance was significantly increased for both effluent treatment types (Fig. 4-15). The magnitude of this effect was quite similar between the two types. There was also a high degree of overlap for taxon richness, although only the activated sludge treatment was significantly greater than zero, when considered as a whole. Thus, the response pattern for these two prevalent treatment types was similar and reflected the mild eutrophication effect that was most commonly observed for benthic invertebrate communities in Cycle 2. The addition of the Gradient mills yielded similar results, except that all the means were pulled in the negative direction, likely due to the inclusion of more marine and estuarine mills (see Section 4.4.2).

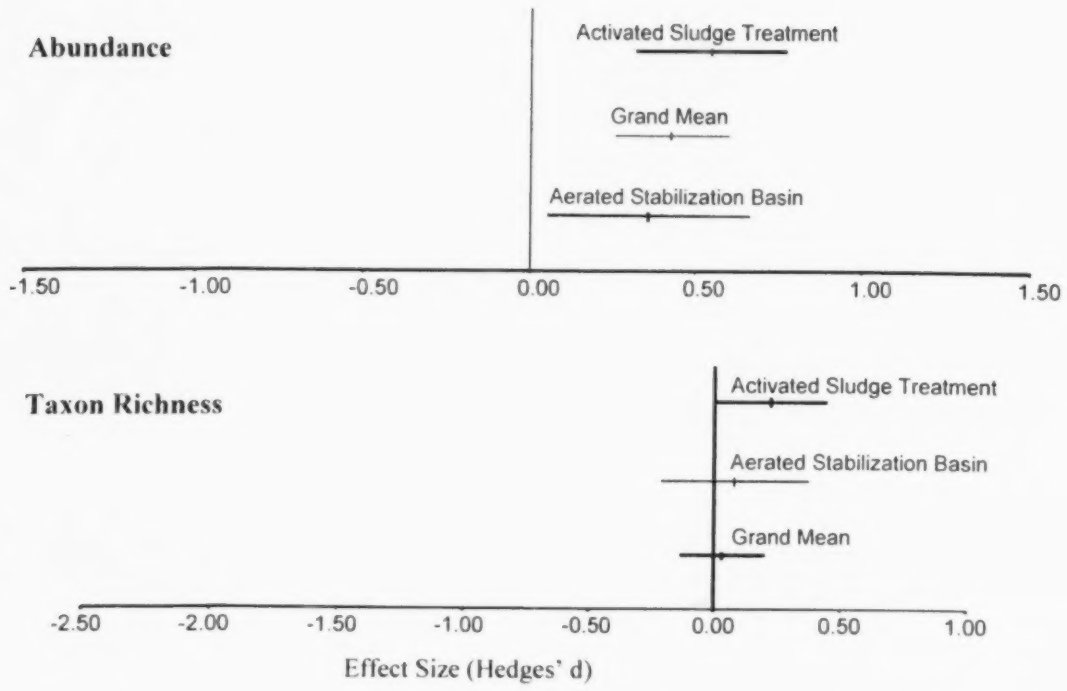


Figure 4-1: Abundance and taxon richness by effluent treatment (Control/Impact mills)

The two major effluent treatment types were also quite similar in terms of effects on evenness (Fig. 4-16). Neither grouping was significantly different from zero. Adding in the Gradient mills moved the means to the right, again probably reflecting the effect seen in marine and, especially, estuarine mills, where effluent exposure may have led to the loss of rare taxa (see Section 4.4.2). As expected, the magnitudes of effects were larger for the Bray-Curtis endpoint (see Section 4.5.1), and these measured changes in community composition were fairly similar between the two treatment types (Fig. 4-17).

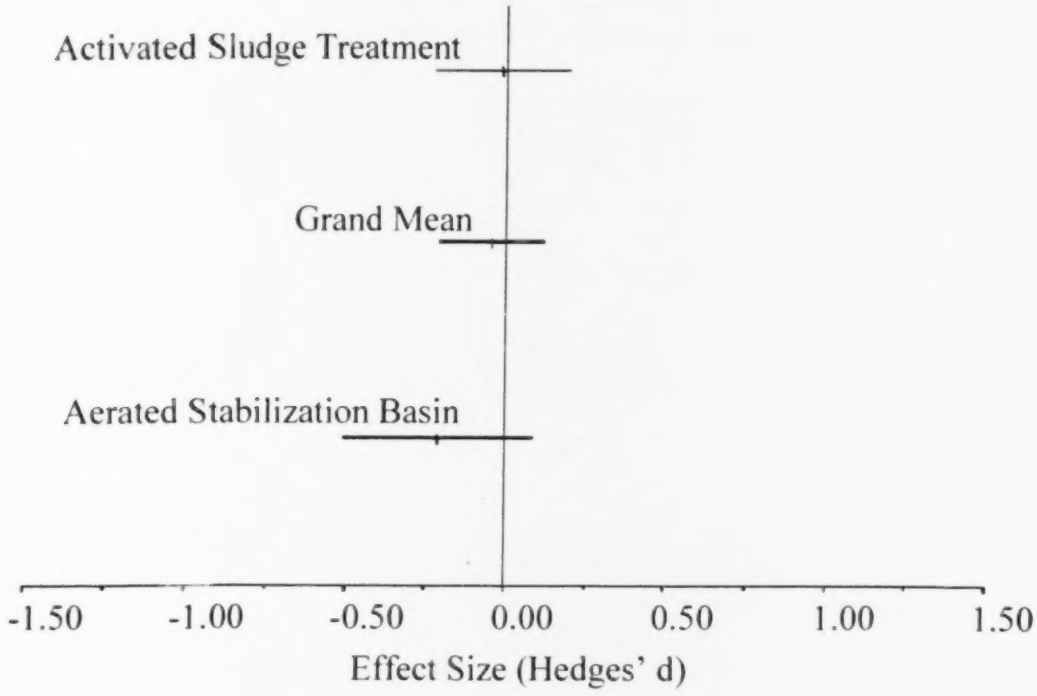


Figure 4-2: Evenness by effluent treatment (Control/Impact mills)

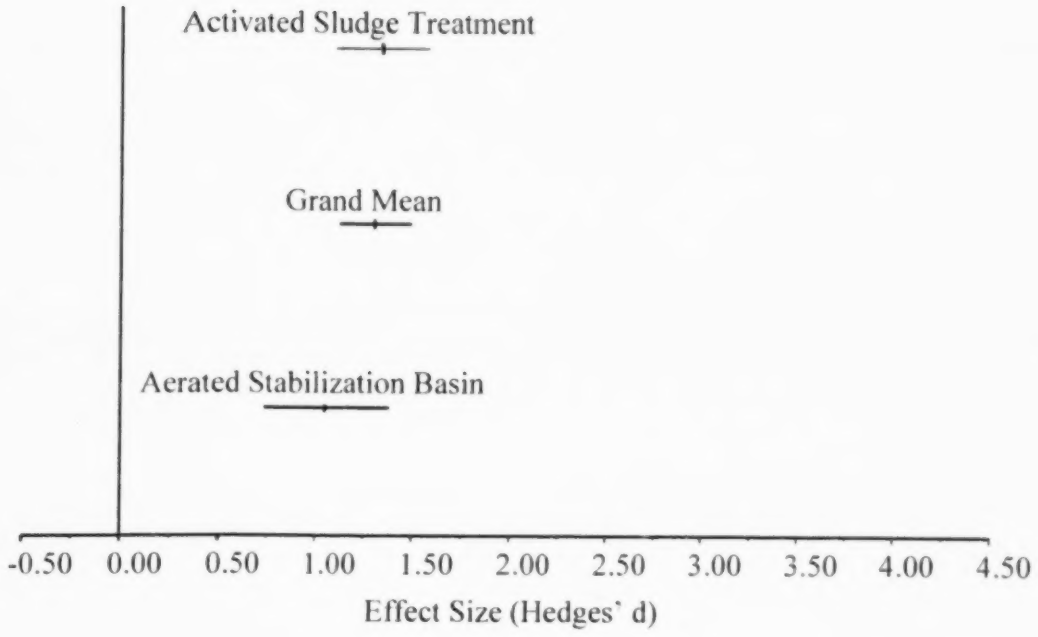


Figure 4-3: Bray-Curtis by effluent treatment (Control/Impact mills)

As was found for effluent treatment, exposure area abundance was also significantly increased to a similar degree for both mill process types (Fig. 4-18). The two types differed for the taxon richness endpoint, though, with the thermomechanical mean being significantly to the left of the kraft mean. This would suggest a milder eutrophication response for kraft mills (more typical of the national norm) and a more pronounced eutrophication response for thermomechanical mills (see Section 4.4.1). A similar pattern was observed when including the Gradient mills. As discussed above, although other co-occurring variables could have been the actual cause of the differences between the two process types, any such co-occurring variables would likely need to have been fairly tightly linked to mill process for this kind of pattern to occur, given the large number of mills in each category.

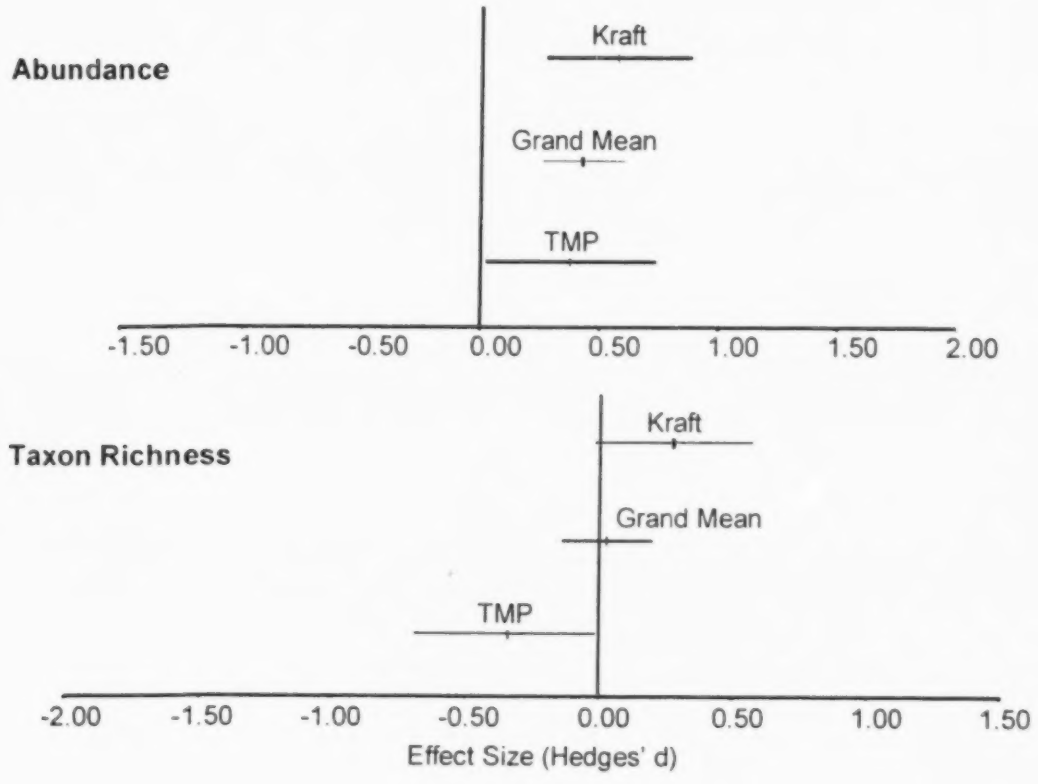


Figure 4-4: Abundance and taxon richness by mill process (Control/Impact mills). Kraft = kraft mills; TMP = thermomechanical pulp mills.

The two mill process type responses were fairly similar for the evenness endpoint (Fig. 4-19). Both 95% confidence intervals overlapped the zero effect size line. This was also true after the addition of the Gradient mills. This indicates that individual mills that had either significant increases or decreases in evenness could not be distinguished on the basis of these two major mill process types. As expected (see Section 4.5.1), the magnitude of effect was greater for the Bray-Curtis endpoint (exposure areas exhibited significant changes in community composition; Fig. 4-20). However, as for evenness, the two mill process types did not significantly differ.

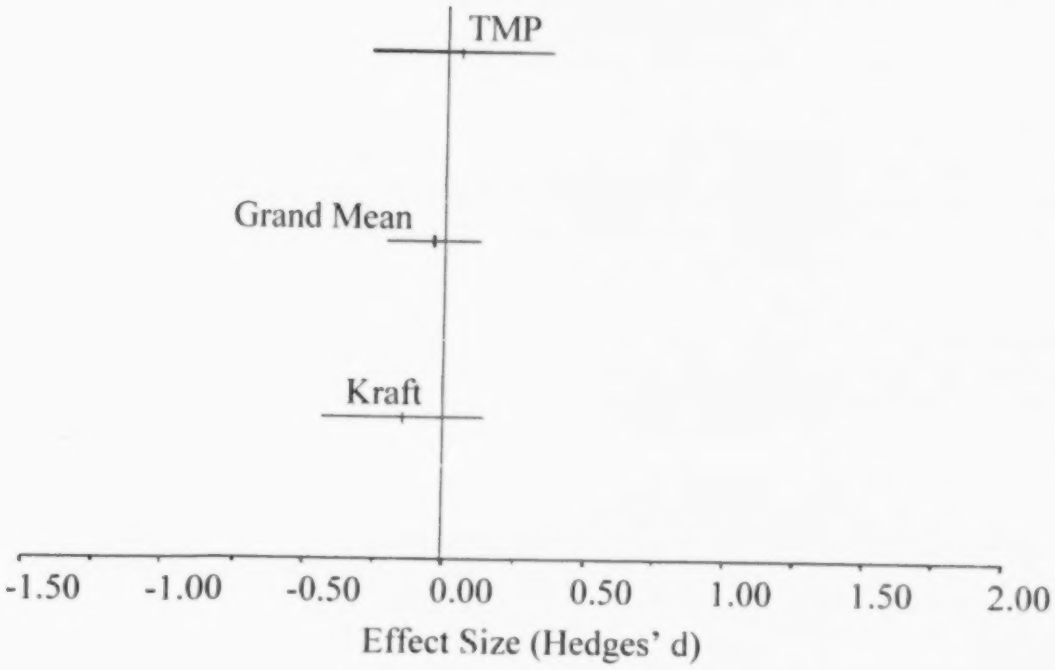


Figure 4-5: Evenness by mill process (Control/Impact mills). Kraft = kraft mills; TMP = thermomechanical pulp mills.

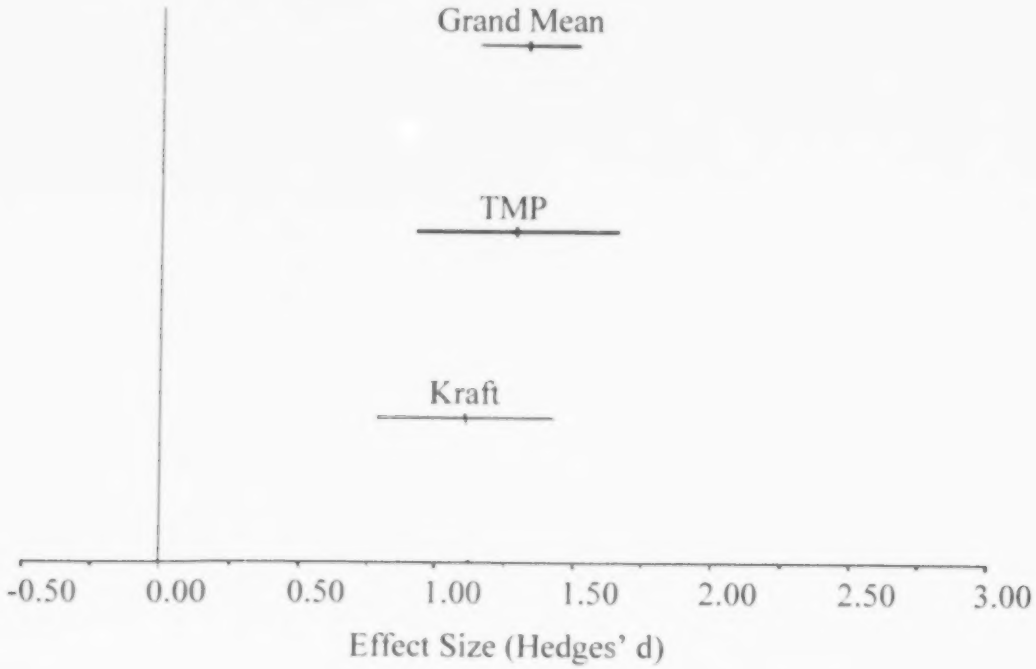


Figure 4-6: Bray-Curtis by mill process (Control/Impact mills). Kraft = kraft mills; TMP = thermomechanical pulp mills.

4.5 Evaluation of Selected Benthic Survey Program Elements

4.5.1 Relative Sensitivity of Endpoints

Each of the core endpoints used in the benthic invertebrate community survey was chosen to help evaluate a different key aspect of community structure that might be altered by exposure to effluent. Because certain aspects of community structure may change more than others following exposure, certain endpoints may, in turn, be more sensitive than others. This does not mean that less sensitive endpoints are less valuable, given that changes in any of the endpoints may occur under certain exposure scenarios. In fact, significant exposure area changes were measured for all four endpoints, as discussed in Section 4.4. Since all four endpoints measure very fundamental community properties, any such changes may have important implications for ecosystem health.

Nevertheless, it is instructive to investigate the relative sensitivity of the four endpoints to determine which aspects of community structure show the greatest changes following exposure. More sensitive endpoints may also have value as an early warning of later, more serious effects. We addressed this question in two ways: 1) tabulation of individual mill results (below) and 2) meta-analyses (Section 4.4).

Using the individual mill ANOVA comparisons (with electronically submitted data) of reference versus exposure areas, we tabulated the number of mills exhibiting a significant increase, decrease, or no significant change in the exposure area for each of the four endpoints (Fig. 4-21). For three of the endpoints (abundance, taxon richness, evenness), the highest percentage of mills fell in the "no significant change" category. This was expected, given that most individual mill studies were designed to have sufficient power to detect only very large effects on the order of two standard deviations. That is, the program, as a whole, was designed to primarily just pick out mills showing the most extreme effects when doing comparisons at the level of the individual mill, which is what is captured for the first three endpoints in Figure 4-21. This issue is discussed more fully in Section 4.5.3. The average effects for these three endpoints were smaller than this (smaller than two standard deviations). Thus, a high percentage of "no significant change" results for the first three endpoints should not be taken as evidence of no effect. Rather, it is evidence that less than half the mills had effects of approximately two standard deviations or more for these three endpoints. Nonetheless, it is possible to look at the average (countrywide) direction(s) of change in Figure 4-21 by noting the relative percentage of mills showing a significant increase or decrease, as follows.

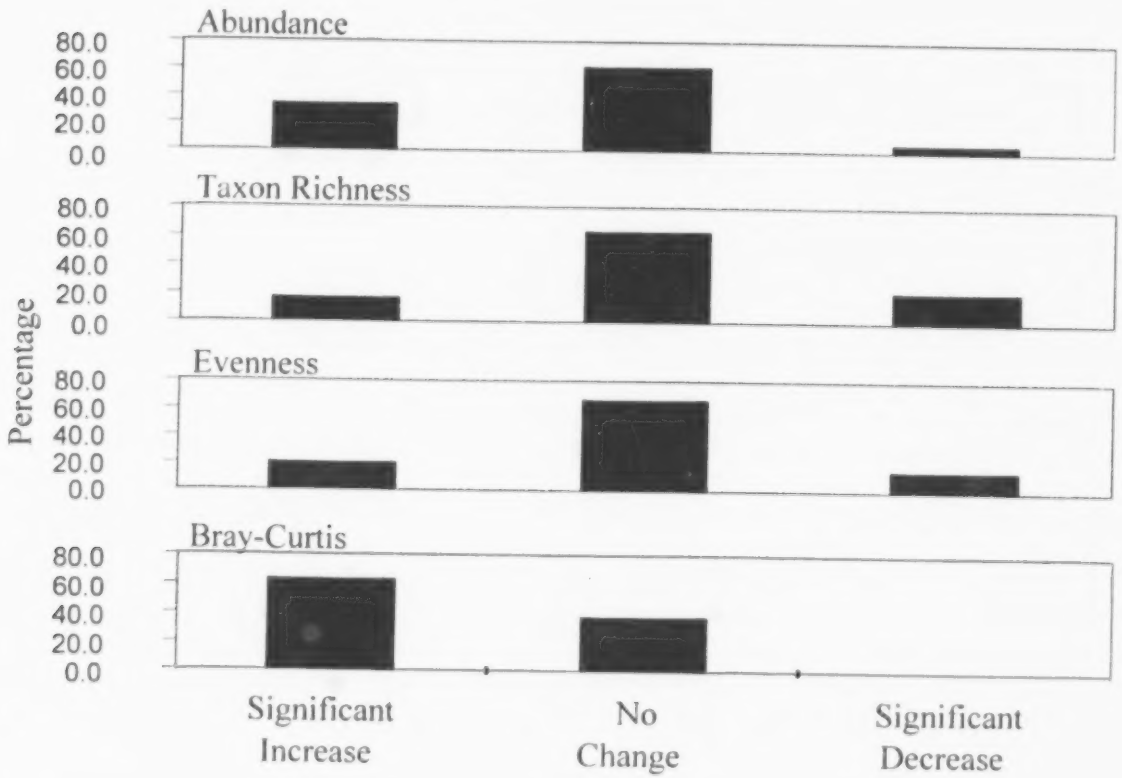


Figure 4-1: Percentage of Control/Impact mills showing significant exposure versus reference differences

For total abundance, there were more significant increases than decreases in the exposure area (Fig. 4-21), reflecting the most commonly observed effect of nutrient enrichment, a pattern that is illustrated more fully in Section 4.4. Taxon richness was more evenly divided between increases and decreases. Increases in taxon richness are expected under conditions of mild eutrophication, while decreases are expected under conditions of more pronounced eutrophication or toxicity/smothering (Fig. 4-3). Evenness was also fairly evenly divided between increases and decreases, and the interpretations of the evenness responses are given in Section 4.4.

In contrast, the Bray-Curtis endpoint showed a quite large percentage of increases in the exposure area, more even than in the "no significant change" category (Fig. 4-21). There were no mills showing a significant decrease in the Bray-Curtis index. This was expected, given the way in which the Bray-Curtis endpoint is calculated (see Section 4.4.1 for further discussion). Thus, this figure shows that over half the mills measured a significant change in overall community structure (particularly community composition) in exposure areas, even given constraints on sampling effort in exposure and reference areas (leading to limited number of replicates). The large percentage of increases is not surprising after inspecting the large average magnitudes of effect shown for this endpoint in the Control/Impact figures in Section 4.4. This illustrates the high sensitivity of the Bray-Curtis endpoint and indicates that a shift in invertebrate community composition is one of the more likely changes to be expected in effluent exposure areas.

Another endpoint sensitivity issue that has been under evaluation in the EEM program is the relative advantage of using diversity versus evenness as an evaluation endpoint. Traditionally, diversity has been used as one of the core endpoints due to its long history of use and its ability to capture changes in both taxon richness (number of taxa) and evenness (how evenly individuals are distributed among the taxa that are present). The fact that diversity is an index that combines both taxon richness and evenness is also a weakness, however. Since taxon richness is already regularly reported as one of the core benthic endpoints, there is little to be gained by capturing it again in the diversity endpoint. Furthermore, since diversity is a hybrid endpoint, its use runs the risk of missing effluent effects on evenness that may be occurring.

This point is illustrated in Figure 4-22, which gives taxon richness, Simpson's diversity, and Simpson's evenness for the Control/Impact marine/estuarine mills. Figure 4-22 shows a particular pattern that was observed for all the mill groupings — the diversity endpoint always fell at some point in between taxon richness and evenness. This is to be expected, given that diversity is a combination of these latter two endpoints. This example shows that reporting diversity (rather than evenness) may yield little information that is not already captured by taxon richness. In contrast, Simpson's evenness has been shown to be independent of taxon richness (Smith and Wilson, 1996). That is, unlike diversity, it measures an important aspect of community structure that is independent of, and in addition to, the information gained from measuring taxon richness. Furthermore, the potential effect of effluent on evenness would have been missed in this example, had evenness not been reported separately. Therefore, of the two, reporting evenness probably has greater value than reporting diversity, as long as taxon richness is also reported.

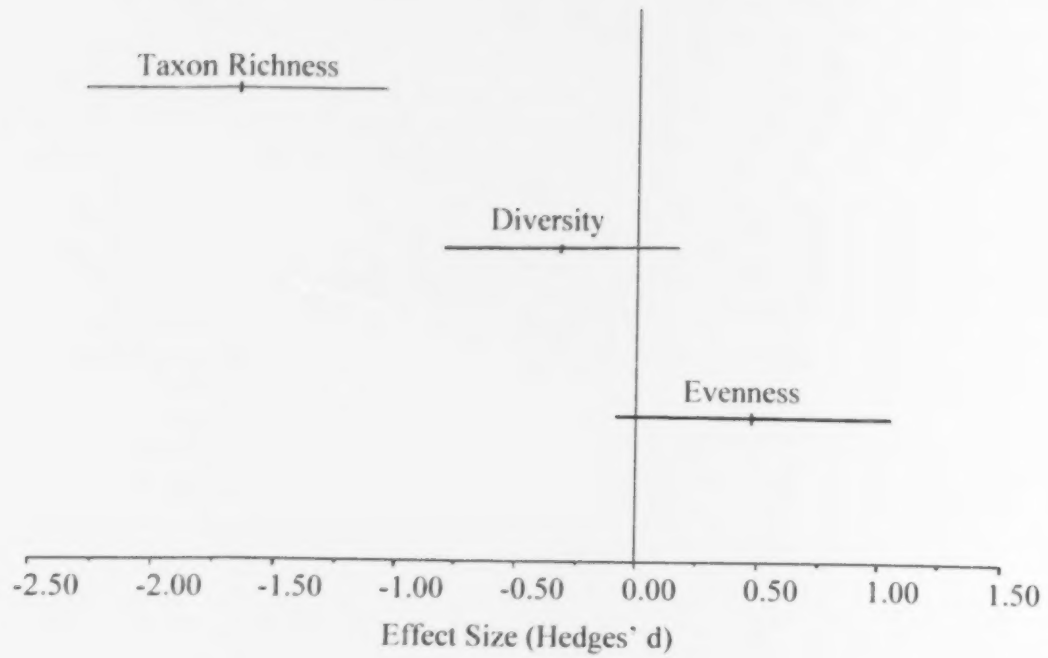


Figure 4-2: Evenness versus diversity comparison (Control/Impact marine/estuarine mills)

4.5.2 Study Design

As discussed above, two basic study designs were used during Cycle 2: Control/Impact and Gradient. Many different permutations of these two designs were used, and these are discussed in the EEM Technical Guidance Document (Environment Canada, 1998b). As part of the national assessment, it was possible to investigate the relative advantages of these two basic designs. In the separate meta-analyses performed on the Control/Impact versus Gradient data sets, it was evident that the different mill groupings were more clearly separated for the Control/Impact designs, even when the same standardized effect sizes were used (Fisher's z-transform).

The relative ability of the two designs to detect effects was also investigated at the level of the individual mill by tabulating the percentage of mills where one, two, three, or no endpoints showed significant differences in reference versus exposure ANOVA comparisons (Control/Impact) or significant correlation coefficients for endpoint versus distance from outfall (Gradient) (Fig. 4-23) (see Section 4.4.1 for a discussion of why the Bray-Curtis endpoint was not included in the Gradient analyses). Figure 4-23 shows that fewer significant differences were detected when using the Gradient design. That is, the Control/Impact design was better able to detect significant effects. A larger proportion of the Gradient design mills discharged into marine/estuarine environments, compared with the Control/Impact mills. However, inspection of

the data suggests that effects in marine/estuarine habitats were likely at least as great as in freshwater habitats. Thus, the reduced resolving power of the Gradient design was probably not simply a consequence of the particular habitat types in which it was used.

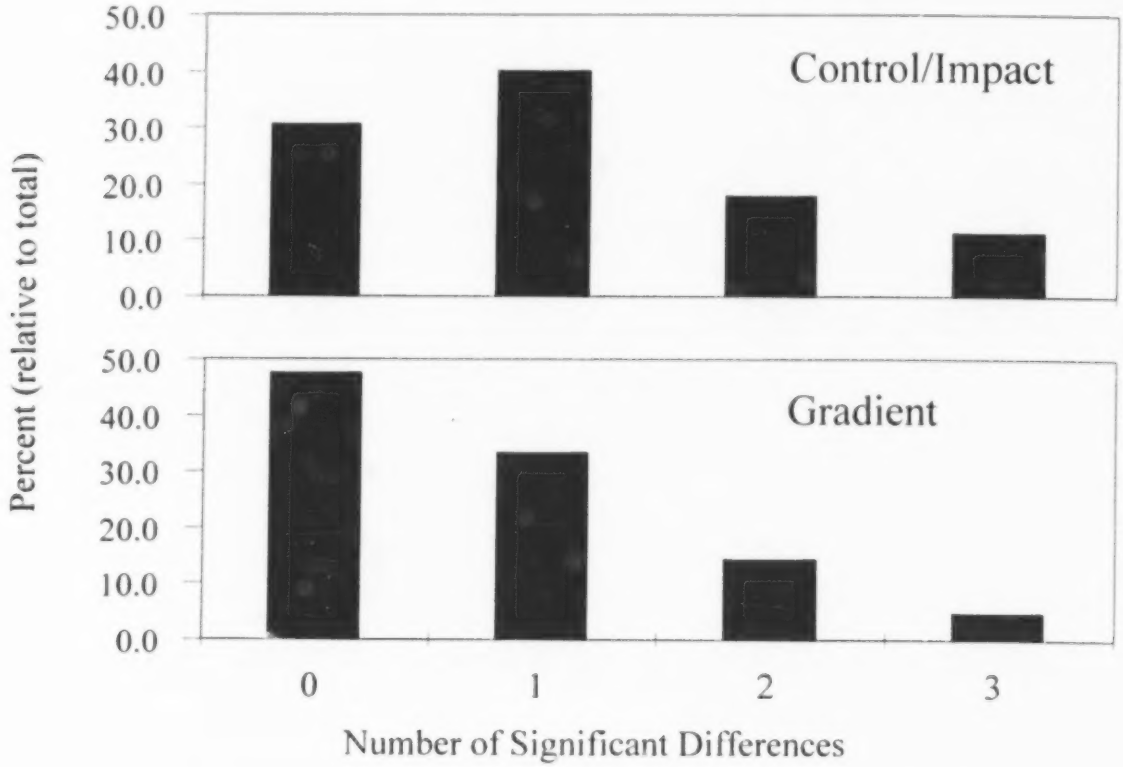


Figure 4-1: Percentage of mills showing significant exposure versus reference differences for zero, one, or a combination of three endpoints (abundance, taxon richness, evenness) for Control/Impact and Gradient mills

The reduced resolving power of the Gradient design was not unexpected, for at least two reasons. The first is related to the patchy distribution of local habitat types in most receiving environments. This makes it difficult to position sampling stations evenly along a gradient without locating some stations in slightly different habitat types (e.g., different grain size, different depth, etc.). It would be expected that sampling in different habitat types would add to background variability in the data. Indeed, many of the regressions of endpoint values versus distance to the outfall were fairly "noisy" (i.e., high variance in the regressions), and this higher variance reduced the resolving power of the Gradient regressions.

The second likely cause of the reduced ability of Gradient designs to identify effects also relates to difficulties in positioning sampling stations. For the Control/Impact designs, near-field exposure areas were compared with reference areas that were entirely unexposed to effluent. For

Gradient designs, however, more exposed stations were essentially being compared with less exposed stations at the far end of the gradient. Therefore, on average, the Gradient designs would not be expected to be measuring the full magnitude of the effluent effects. It would be possible to span most of the range from near-field exposure to the full reference condition in a Gradient design if the least exposed sampling stations are positioned far enough from the outfall to approximate the reference condition. However, this was often not possible due to constraints in available sampling areas.

On the other hand, the Gradient design also has certain advantages. It can be better at detecting spatial patterns in effects, because sampling stations are spread over a greater range of locations than in the Control/Impact design. This can be helpful in determining how the nature of effects changes with increasing distance from the outfall, some of which may be due to changes in habitat type. Gradient designs are also helpful in estimating the geographic extent of effects by judging where along the gradient effects are asymptotic to the reference condition. In addition, Gradient designs can be useful when rapid effluent dilution precludes extensive sampling within an exposure area with a relatively homogeneous effluent concentration. It has been suggested that Gradient designs are the only option when a true reference area is not available. However, this is not true, given that Control/Impact designs can also be used in this situation. In this case, the reference stations in a Control/Impact design would be relocated to the equivalent of the end of the gradient (farthest available comparable habitat from the effluent) so as to estimate effluent effects in effectively the same way as for a Gradient design.

These findings suggest a few key ways in which to improve Gradient designs in future cycles, when conditions allow. First, background noise is usually reduced and, therefore, resolving power increased when all sampling stations are located in comparable local habitat types. This may require that stations are not spaced evenly along the gradient. Second, when feasible, the gradient should extend far enough from the outfall so that the most distant sampling stations approximate the reference condition as much as possible. Third, greater sampling effort (greater number of sampling stations) should be expended on the two ends of the gradient. This would both increase the resolving power of the regression analyses and provide data appropriate for concurrent ANOVA comparisons; the latter would provide results more directly comparable with those obtained from Control/Impact designs.

4.5.3 Effect Size for Calculating Required Sampling Effort

The determination of appropriate predetermined effect sizes is an issue that has a number of important implications for the EEM program. For example, predetermined effect sizes have the potential for being quite useful for evaluating the ecological significance of measured effects. To illustrate, if it is known that a change in taxon richness of over 40% would have long-lasting, deleterious effects on the functioning of the receiving water ecosystem, then a measured change in taxon richness of 50% in a particular exposure area could be viewed as an effect of sufficient magnitude to raise some concern. In this example, 40% would be used as a predetermined critical effect size, exceedence of which might require follow-up action. Unfortunately, most receiving water ecosystems in the EEM program are not yet well enough understood to set these kinds of critical effect sizes.

Nevertheless, initial efforts are under way to establish predetermined effect sizes for the purpose of calculating required sampling effort. This second kind of predetermined effect size is related to, but not the same as, a critical effect size. Rather, it is the magnitude of effect that the EEM program would like to be able to detect. It may be partly based on initial ideas of what would be a critical effect size, but the two will not necessarily be equal. That is, it may later be decided that greater or lesser effects are of interest with respect to taking corrective action. For example, expanding on the previous example, the predetermined effect size for taxon richness may be set at ±40% at the beginning of a study for the purposes of calculating required sampling effort, based on initial ideas of what would be an ecologically (or other) important effect. Following the study, however, it may be decided that a 30% effect is unacceptable due to priorities set, for example, by local stakeholders (other interested parties may have similar input). Thus, the predetermined effect size used for calculating sampling effort may be related to, but does not necessarily equal, a critical effect size used for making decisions regarding the need for corrective action.

In addition to these two kinds of predetermined effect sizes, a third kind of effect size is the magnitude of effect that is actually measured in a given study (e.g., the 50% example in the previous paragraph). This measured percent change can be relative to the reference mean (as in Fig. 4-24 below) or relative to background variability (as in the Hedges' d measure used in some of the above meta-analyses).

The need for the second kind of predetermined effect size (for calculating sampling effort) becomes apparent when considering the standardized power analysis equation (below). This equation is used to calculate the number of samples (n) that need to be collected in a particular study to achieve a desired level of statistical power (1-β). Statistical power and the following equation are discussed in more detail in Lowell (1999) and Lowell et al. (2002), but, basically, statistical power is related to the ability of a particular study to detect an effect of a given magnitude, if it has occurred. The goal is to increase power by increasing sampling effort (n), within the constraints of the monetary cost of increasing sampling effort. To calculate required sampling effort, it is first necessary to set a predetermined effect size (ES) for use in the following equation (Green, 1989):

$$n = 2(t_\alpha + t_\beta)^2(SD/ES)^2$$

where:

- n = sample size
- t_α = t value for α significance level (Type I error; two-tailed)
- t_β = t value for β significance level (Type II error; one-tailed)
- SD = standard deviation
- ES = predetermined effect size

One of the proposals under consideration for the benthic invertebrate component of the EEM program has been to initially set ES in terms of the degree of departure from the normal range of variability observed in reference areas. One measure of the normal range of variability that has

been used in the literature is ±2 standard deviations about the mean for the reference community (Lowell, 1997).

The relationship between measured effect size (absolute value) and measured two standard deviations (2SD) for each of the core endpoints is given in Figure 4-24. These are given as the medians, across the country, of the values measured at each mill during Cycle 2 (including mills where the reference and exposure areas were not significantly different). Thus, this figure shows how big the measured effects were, on average, relative to what might be termed the normal range of variability. For total abundance, the median change in exposure areas (relative to reference areas) was approximately 70%, although the variability among mills was also quite high for this endpoint. The measured effect sizes for the other endpoints ranged from about 20% to 90%.

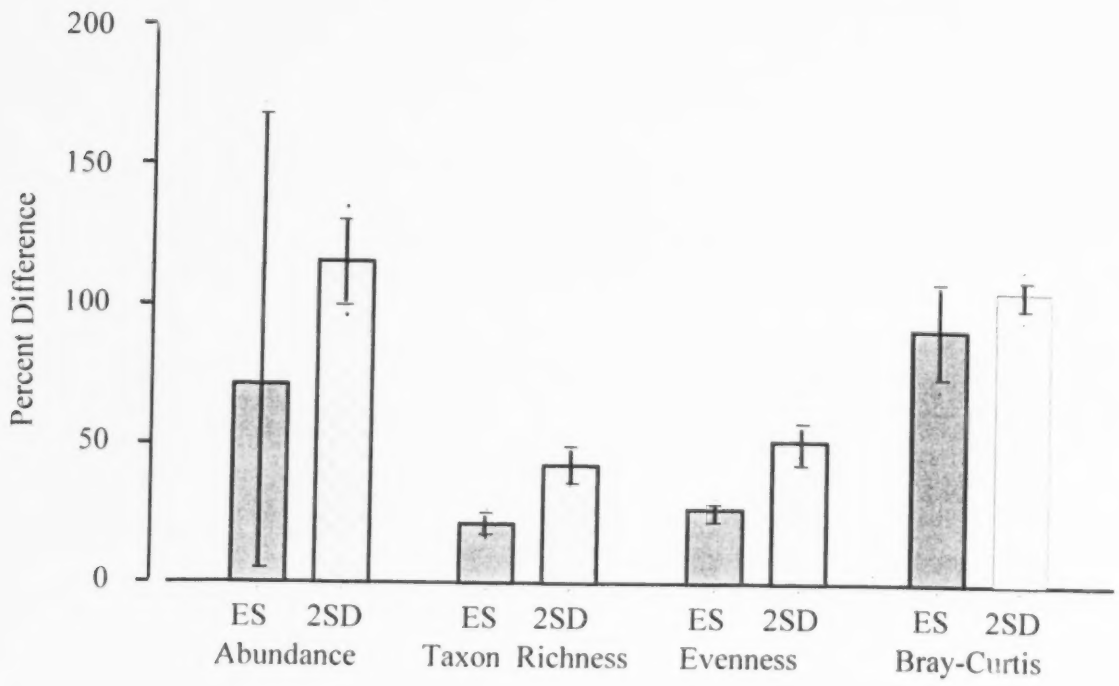


Figure 4-1: Measured effect size (ES) and "normal range of variability (2SD)" (relative to reference areas)

Thus, despite pronounced improvements in pulp mill effluent quality over the last decade (see discussion of the sublethal toxicity data in Chapter 7), the magnitude of change measured in exposure areas remains fairly high. On the other hand, on average, the measured effects were roughly half the normal range of variability seen in reference areas, if that is defined as two standard deviations. The exception was the Bray-Curtis endpoint, indicating that changes in

community composition in exposure areas were greater, relative to background variability, than the changes measured by the other three endpoints. This helps explain why the Bray-Curtis measure was the most sensitive measure of the four core endpoints (see Sections 4.4.1 and 4.5.1).

The value of using 2SD as a predetermined effect size (for calculating required sampling effort) can be seen by calculating what percentage of the individual mill exposure areas had measured effect sizes that exceeded 2SD for their respective reference areas — that is, what percentage of the measured effects exceeded what might be termed the normal range of variability. These percent exceedences ranged from roughly 20% to 45% (Figure 4-25).

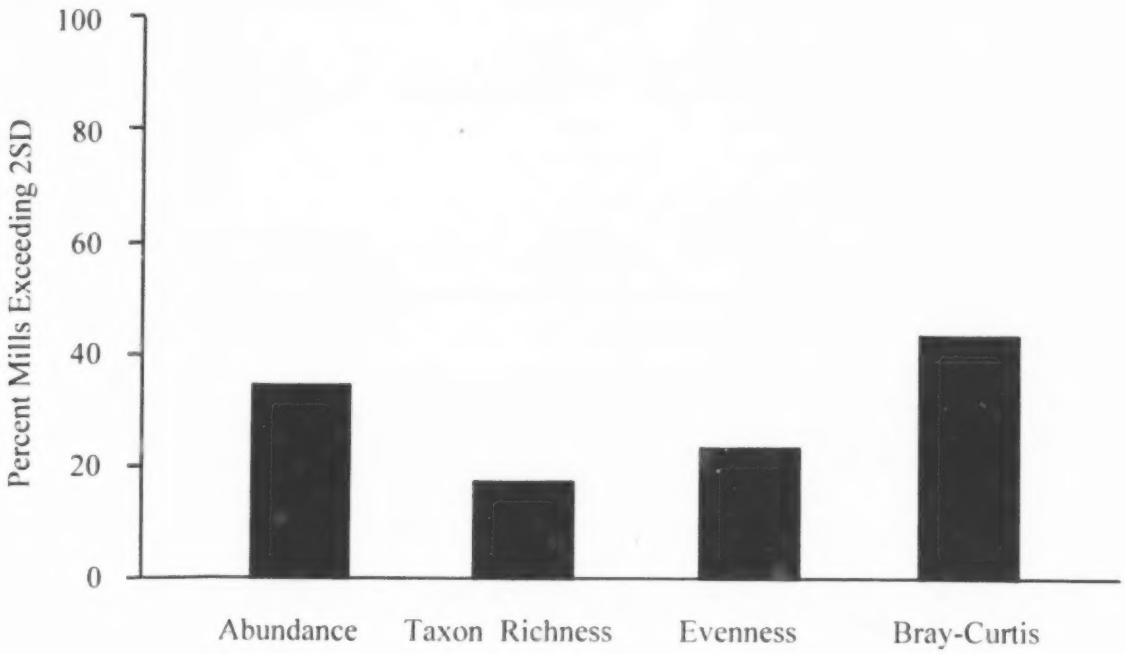


Figure 4-2: Percentage of mills where measured effect size exceeded "normal range of variability (2SD)"

This helps to give an estimate of how useful 2SD might be for distinguishing more heavily impacted sites from less heavily impacted sites. If the predetermined effect size was being set too low, then most of the measured effects would exceed the predetermined effect size, and the percentages in Figure 4.25 would tend towards 100%. If the predetermined effect size was being set too high, then few of the measured effects would exceed the predetermined effect size, and the percentages in Figure 4.25 would tend towards 0%. The fact that the percentages were at a more intermediate level suggests that using 2SD as a predetermined effect size has some value for determining which mill exposure areas are showing the most pronounced effects. It should be further noted, however, that it may be necessary to use other predetermined effect sizes in future

stages of the EEM program, as more information becomes available on the magnitudes of effects that are of concern.

4.6 Conclusions

- Effects of pulp mill effluent on benthic invertebrate communities appear to have improved since Cycle 1.
- Effects are still widely present, however, they do not appear to be pronounced at most mills.
- The national average response pattern for benthos was one of mild to moderate eutrophication.
- The national average response pattern for riverine benthos was likewise one of mild to moderate eutrophication.
- A pronounced eutrophication response pattern was also evident at a smaller number of mills.
- The national average response pattern for marine benthos was one of toxicity/smothering.
- Lacustrine and estuarine benthic effects were intermediate between riverine and marine benthic effects.
- Although the above response patterns were those observed most frequently, individual mill receiving environments sometimes showed response patterns that differed from the national average for a particular habitat type.
- Except for the Bray-Curtis endpoint (change in community composition), the magnitudes of effects were, for the most part, less than two standard deviations. The ecological significance of these magnitudes remains to be determined.

5. Fish survey

5.1 Introduction

One of the principal components of the EEM program is the fish survey, which monitors fish in the receiving water where mills discharge effluent. The fish survey is used to determine if the effluent is modifying fish populations, impacting the ability of Canadians to use the fish and protect the safety of consumers

Tainting tests, designed to determine if the taste of the fish is affected, and analyses for dioxins and furans in edible portions of fish are conducted to determine if the effluent is impacting the ability of Canadians to use the fish and to protect the safety of the consumers. Only two mills conducted a tainting study as part of Cycle 2. In one case the mill was identified as the cause of the taint, and measures are currently under way to mitigate this problem. In the other case, the tainting evaluation confirmed that tainting was an issue in the exposure area but the cause could not be identified.

In Cycle 1, mills that were using chlorine bleaching or had used chlorine bleaching in the past were required to conduct a tissue analysis of edible portions of fish for chlorinated dioxin and furan congeners. Where results were below fish consumption guidelines in these evaluations the mill was exempted from these analyses in future cycles. Cycle 1 results indicated that levels of dioxins and furans in edible fish tissue were low. As a result, in Cycle 2, only ten mills conducted fish tissue analyses for dioxins and furans, six of which had levels of dioxins and furans in fish tissue that exceeded Health Canada fish consumption guidelines. Continued monitoring is needed at these six sites so that fish consumption advisories can be annually updated, and with time eliminated, to allow unrestricted consumption of fish at these locations. These results indicate that impacts on fish usability from pulp and paper effluent are localized and not of national concern.

The approach for the fish population survey is based on a comparison of measurements of growth, reproduction, condition, and survival of fish species exposed to effluent with those taken from the same fish species collected in a reference area. The basic assumption with this type of fish population survey is that by looking at the health of one or two fish species, information about the fish community and possibly other aspects of the aquatic ecosystem can then be inferred (EC/DFO, 1993). This method provides a cost-effective method of determining where fisheries concerns exist (Munkittrick et al., 2002).

Originally, the purpose of the fish population survey was to assess the health of fish exposed to mill effluent (EC/DFO, 1993); however, this objective was modified prior to the completion of the first cycle. The purpose of Cycle 1 was to use the monitoring information collected to determine the variability in fish population survey measurements (Environment Canada, 1997c). Cycle 1 was also meant to enable the evaluation of various aspects of EEM study designs, such as the selection of fish species, sampling areas, and capture techniques. The objective of the second and subsequent EEM cycles was then to determine the presence of effects and to identify causative factors associated with them (EC/DFO, 1993).
The following core endpoints were used to assess potential effects of effluent on fish:

- gonad weight (indicator of reproduction);
- condition factor (indicator of condition);
- liver weight (indicator of condition);
- weight at age (indicator of growth); and
- age (indicator of survival or other effects on age structure).

This chapter provides a national evaluation and synthesis of the type and magnitude of effects of pulp and paper mill effluents on fish in Canadian receiving waters. Furthermore, as outlined in Chapter 1, emphasis is placed on certain priority areas identified by the National EEM Science Committee, including the influence of gender, species, mill process type, effluent treatment type, and habitat type on measured effects, as well as the evaluation of selected fish population survey elements. The assessment also provides the basis for recommendations on how to improve the EEM program for future cycles.

Two main approaches were used to look at Cycle 2 fish population survey data: 1) tabulation of the results of the individual mill ANCOVA/ANOVA analyses and 2) meta-analyses of the magnitudes of effects (see Chapter 3). A preliminary examination of the Cycle 2 fish population survey data was carried out on a mill-by-mill basis, using the ANCOVA results that were submitted by mills in their interpretative reports (Munkittrick et al., 2002). Following this initial analysis, the Cycle 2 electronic data submitted by the mills to Environment Canada underwent a quality assurance/quality control process, resulting in the removal of some data points. The electronic data set was then used to conduct ANCOVA comparisons similar to those reported in Munkittrick et al. (2002), which provided information on:

- the number of statistical differences;
- the consistency of differences that were found (e.g., was the same effect present in the same direction in both sexes?);
- the magnitude of changes in a specific endpoint; and
- the consistency between data reported by mills in their interpretative reports and their electronic data.

The electronic data also provided valuable information on the types of data entry errors, which was subsequently used to improve the electronic reporting system.

The electronic data were also used to conduct meta-analyses. For the fish population survey, Cycle 2 data were analyzed using meta-analysis to determine and compare the magnitude of effects on the key fish endpoints (gonad weight, liver weight, condition, weight at age, and age) as influenced by sex of the fish, the fish species selected, the mill process used, the effluent treatment type, and the habitat of the receiving area.

5.2 Improvement in the Fish population survey — Cycle 1 versus Cycle 2

Cycle 1 of the pulp and paper EEM program was completed in 1996. As discussed in previous chapters, there were a number of problems associated with monitoring during Cycle 1. For example, only 8% of mills (9 of 115 studies) successfully caught the required minimum number of fish (at least 20 fish of each sex of two fish species) in the exposure area and reference area (Environment Canada, 1997c). Fifty percent of the mills were marginally or partially successful (e.g., one or two fish species with at least 15 fish of each sex captured), and 42% were unsuccessful in capturing fish (not enough fish of either species or sex). Table 5-1 summarizes the typical problems encountered in Cycle 1.

Table 5-1: The number of mills that encountered various problems in the Cycle 1 fish population survey within the freshwater receiving environment (from Environment Canada, 1997c)

Problems or Confounding Factors	No. of Mills — River (n = 73)	No. of Mills — Lake (n = 10)
Too few fish	42	7
Uncertain exposure	23	2
Reference sites	16	1
Other effluents	15	3
Only one fish species	13	3
Incomplete/poor reporting	12	3
Gonads undeveloped/spent	10	0
Fish immature	9	3
Physical confounding factors (e.g., shipping, fast current)	6	0
Aging problems/not aged	6	0

Cycle 2 EEM surveys were conducted between 1997 and 2000, with the mill interpretative reports being submitted to Environment Canada by April 1, 2000. In general, fish population surveys were more successful during Cycle 2; 114 interpretative reports contained fish population survey information. Over 56% of mills (64 of 114 interpretative reports) attempted a standard fish population survey (e.g., capturing 20 males and 20 females of two fish species), and the majority of these mills were successful in obtaining interpretable data. In Cycle 2, only one mill that attempted a fish population survey was not successful in obtaining at least 15 individuals of at least one species of fish. In Cycle 2, over 82% of the studies that attempted to use fish were successful in getting sufficient information for interpretation, whereas well under half of the studies in Cycle 1 actually reported interpretable data.

The most common freshwater finfish species collected in Cycle 1 were the white sucker (n = 46), longnose sucker (n = 24), walleye (n = 10), and shorthead redhorse sucker (n = 8) (Table 5-2). Most of the concerns regarding the use of these fish species related to their range of movement, which could make exposure of fish to effluent questionable. For marine/estuarine studies, the most common species used were the winter flounder (n = 7), English sole (n = 5), tomcod (n =

5), and dungeness crab (n = 6). Concerns with the marine/estuarine species were mainly with the difficulty in capturing sexually mature species, migratory patterns, and aging.

The most common freshwater finfish species used for Cycle 2 monitoring, as reported in the interpretative reports, were the white sucker (32 studies) and longnose sucker (20 studies), followed by the yellow perch (9 studies) and johnny darter (6 studies) (Munkittrick et al., 2002). Fourteen species that were not used in Cycle 1 were chosen for monitoring in Cycle 2. All of these species, such as minnows, darters, and sculpins, are considered small-bodied species. Table 5-2 describes the freshwater and marine fish species used in Cycle 1 and Cycle 2.

Table 5-2: Species used (in order of frequency in Cycle 1 and Cycle 2) at freshwater and marine/estuarine sites

Species	Scientific Name	Cycle 1	Cycle 2	
Freshwater				
White sucker	*Catostomus commersoni*	46	32	
Longnose sucker	*Catostomus catostomus*	24	20	
Walleye	*Stizostedion vitreum*	10	1	
Yellow perch	*Perca flavescens*	7	9	
Shorthead redhorse sucker	*Moxostoma macrolepidotum*	8	1	
Mountain whitefish	*Prosopium williamsoni*	5	1	
Brown bullhead	*Ameiurus nebulosus*	4	3	
Largescale sucker	*Catostomus macrocheilus*	4	3	
Silver redhorse sucker	*Moxostoma anisurum*	4	3	
Spottail shiner	*Notropis hudsonius*	4	2	*
Common carp	*Cyprinus carpio*	2	1	
Lake chub	*Couesius plumbeus*	2	1	*
Peamouth chub	*Mylocheilus caurinus*	2		
Rock bass	*Ambloplites rupestris*	2	2	
Burbot	*Lota lota*	1	1	
Channel catfish	*Ictalurus punctatus*	1	1	
Fathead minnow	*Pimephales promelas*	1		*
Goldeye	*Hiodon alosoides*	1		
Northern pike	*Exos lucius*	1		
Threespine stickleback	*Gasterosteus aculeatus*	1		
Smallmouth bass	*Micropterus dolomieu*	1	1	
Lake whitefish	*Coregonus clupeaformis*	1		
Pumpkinseed sunfish	*Lepomis gibbosus*	1	1	
Longnose dace	*Rhinichthys cataractae*		5	*
Trout-perch	*Percopsis omiscomaycus*		4	*
Johnny darter	*Etheostoma nigrum*		6	*
Iowa darter	*Etheostoma exile*		2	*
American chub	*Semotilus corporalis*		2	*
Logperch	*Percina caproides*		1	*
Bluntnose minnow	*Pimephales notatus*		1	*
Blacknose dace	*Rhinichthys atratulus*		1	*

Species	Scientific Name	Cycle 1	Cycle 2	
Common shiner	*Notropis cornutus*		1	*
Redside shiner	*Richardsonius balteatus*		1	*
Mimic shiner	*Notropis volucellus*		1	*
Mottled sculpin	*Cottus bairdii*		1	*
Slimy sculpin	*Cottus cognatus*		1	*
Spoonhead sculpin	*Cottus ricei*		1	*
Marine/estuarine				
Winter flounder	*Pleuronectes americanus*	7	2	
English sole	*Parophrys vetula*	5		
Tomcod	*Microgadus tomcod*	5		
Starry flounder	*Platichthys stellatus*	2		
Longhorn sculpin	*Myoxocephalus octodecemspinosus*	2	2	
Prickly sculpin	*Cottus asper*	2		
Threespine stickleback	*Gasterosteus aculeatus*	2	3	*
Cunner	*Tautogolabrus adspersus*	1		
Mummichog	*Fundulus heteroclitus*	1	3	*
Rock sole	*Lepidopsetta bilineata*	1		
Quillback rockfish	*Sebastes maliger*	1		
Atlantic silverside	*Menidia menidia*	1		*
Slender sole	*Lyopsetta exilis*	1		
Widow rockfish	*Sebastes entomelas*	1		
Shorthorn sculpin	*Myoxocephalus scorpius*		2	
Dungeness crab	*Cancer magister*	6		**
Rock crab	*Cancer irroratus*	3		**
Oyster	*Crassostrea gigas*	2	2	**
Blue mussel	*Mytilus edulis*	2	3	**
Green crab	*Carcinus maenus*	1		**
Macoma	*Macoma balthica*		1	**
Waved whelk	*Buccinum undatum*		1	**

Notes:

1. The data in this table include information from both successful and unsuccessful studies (modified from Environment Canada, 1997d; Courtenay et al., 2002; Munkittrick et al., 2002).
2. Species marked with an asterisk are small-bodied species. Species marked with two asterisks represent species that are not finfish, such as crabs and molluscs.

There was greater success in capturing finfish at marine mills in Cycle 2, compared with Cycle 1, mainly because of the use of small-bodied fish in Cycle 2 (e.g., mummichog). Only one mill of the seven marine/estuarine mills that attempted to capture finfish failed to capture enough fish for data analysis. Four surveys caught enough fish of one species, and two surveys caught enough fish of two species. Three marine mills chose small-bodied fish species — mummichogs and threespine sticklebacks. Mummichogs, in particular, were caught in sufficient numbers, and all EEM endpoints could be measured. There was less success with the use of the threespine stickleback, due to problems with aging the fish.

In Cycle 2, fewer than half of the marine pulp mills (12 of 28) attempted a standard fish population survey, compared with almost 70% of the freshwater pulp mills. Of the 12 marine mills that attempted a standard fish population survey, only seven targeted a finfish species, while the other five used molluscs for the fish population survey (Table 5-2). A greater number of mills made use of alternative methodologies during Cycle 2 to alleviate difficulties encountered with the standard fish population survey in marine/estuarine environments in Cycle 1. Six marine mills and 10 freshwater mills used an alternative monitoring method, conducting a caged bivalve or a mesocosm study (Courtenay et al., 2002; Munkittrick et al., 2002).

5.3 Data Extraction and Preparation for Analysis

5.3.1 Data Extraction and Preparation

As a preliminary evaluation of pulp mill effects seen in Cycle 2, Munkittrick et al. (2002) extracted information from the reports submitted by the mills to Environment Canada. They found that, based on the reports, the most consistent information reported was related to gonad weight (energy use), condition factor ($k = 100*\text{weight}/(\text{length})^3$), and liver weight (energy storage). Information related to age and weight at age was not included in Munkittrick's analysis because of poor/inconsistent reporting.

As previously indicated, data were also submitted electronically by the mills to Environment Canada (MS Excel spreadsheet), along with the written reports. The data were then uploaded into an MS Access database by the National EEM Office.

The electronic data were analyzed via ANOVA (age) or ANCOVA (remaining four endpoints), which provided exposure versus reference area comparisons for each mill/species/sex combination. For the ANCOVA comparisons, liver and gonad weights were regressed against body weight, body weight was regressed against length (condition), and body weight was regressed against age (weight at age). Dependent variables and covariates were $\log_{10}$ transformed to improve linearity and to make variances more homogeneous. SYSTAT 10™ software was used to conduct the analyses. Data were included in the analysis if 15 or more fish by sex were collected at both the reference area and exposure area. Capture success and species selection were also tabulated. The variables used during the statistical analysis are summarized in Table 5-3.

Table 5-1: Endpoints used to determine effects in EEM, as determined by statistically significant differences between exposure and reference areas

Test	Attribute	Dependent Variable	Covariate
ANCOVA	Weight at age	Body weight	Age
	Gonad weight at body weight	Gonad weight	Body weight
	Condition (weight at length)	Body weight	Length
	Liver weight at body weight	Liver weight	Body weight
ANOVA		Age	

Using the electronic data, meta-analyses were also conducted on distinct combinations of mill processing type, treatment type, habitat type, fish species, and gender.

5.3.2 Discussion of Data Entry Errors

Kilgour (2001) examined the Cycle 2 electronic data from the national database for quality assurance/quality control purposes. The data were plotted on a mill-by-mill basis in order to conduct the initial data screening. The scatterplots were inspected to identify outliers and other unusual data. These data were removed in situations where the values were not biologically possible. As well, the visual examination of the data allowed for the identification of immature fish that were included in the submission of the electronic data. In situations where suspected immature fish were identified, but where there were few immatures and even fewer found in both the exposure and reference areas, data were analyzed twice — once with the immatures and once without. In most cases, immatures were excluded from the final analyses. Data that showed a lack of variability in organ weights were also excluded from the analyses when it appeared that the balances were not sensitive enough for the size of the fish.

Other problems that were identified in the electronic data submissions included mills having insufficient reference area data or insufficient exposure area data, as well as the resubmission of electronic data from Cycle 1 (rather than Cycle 2) (Table 5-4).

Table 5-1: Data entry and analysis problems found in the review of the electronic data submitted for Cycle 2 EEM reports (from Kilgour, 2001)

Problems and Confounding Factors	Number of Problems
Too few fish	28
Used immature fish	17
Outliers included in data analysis	9
Fish were not aged	6
Balances used not sensitive enough	5
Fish not caught at reference site	5
Data entry errors	3
Wrong data reported	1
Fish not caught at exposure site	1

Four mills that had used multiple reference areas submitted electronic data. These data were examined to determine if the reference data could be pooled and analyzed as one reference area. Subsequently, two analyses were carried out — once pooling the reference stations and once using only the reference station that was closest to the mill outfall. Only the latter, single-reference area comparisons were used in the final analyses.

5.4 National Trends — ANCOVA

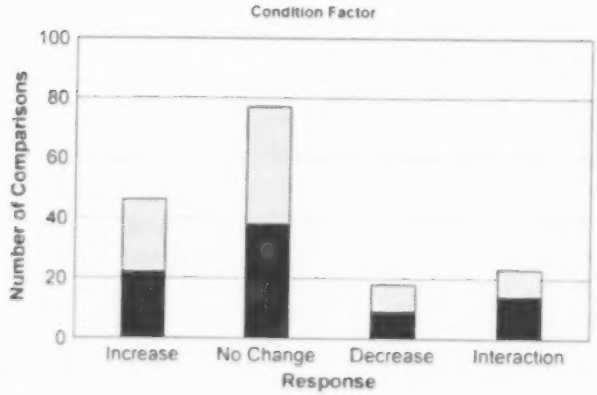
A summary of the ANCOVA results, using both the data that were extracted from the mill interpretative reports (Munkittrick et al., 2002) and the electronic data, was created to facilitate comparisons among mills at a national level. The following comparisons focus on three key endpoints — gonad weight, liver weight, and condition factor (k). A major difficulty encountered when summarizing the mill-submitted reports was the extraction of information required for the analyses, due to the different reporting formats and styles. In most cases, scatterplots of the data were not included in the reports.

Munkittrick et al. (2002) analyzed data from 53 EEM studies, resulting in a total of 172 ANCOVA comparisons. In this study, the requirement was that more than 15 fish of each sex be collected at both the exposure area and the reference area in order to be included in the analysis. Electronic data were submitted for 62 EEM studies; after quality assurance/quality control of the data, 163 ANCOVA comparisons were made. In the following figures and discussion of the ANCOVA results, an "interaction" refers to the case where the reference and exposure area slopes were significantly different. As discussed in Section 5.4.2, a significant interaction is interpreted as an effect related to exposure to the effluent. It was found that there was generally good agreement between the results that were submitted by mills and the electronic data.

5.4.1 Results from ANCOVA

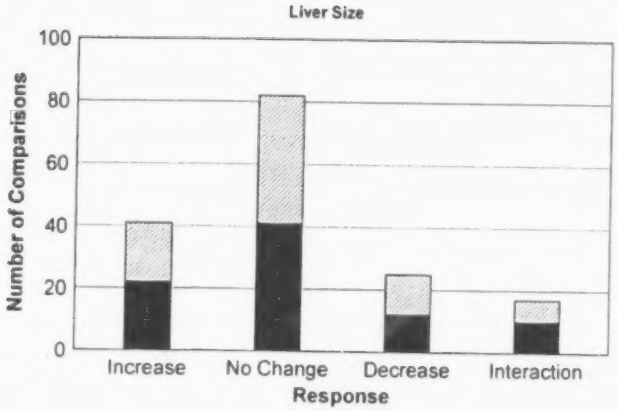
The incidence of effects was consistent between male and female fish. This indicated that in a high percentage of comparisons, when an effect was found in female fish, the same effect (in the same direction) was found in male fish of the same species (Fig. 5-1). Similar results were found between fish species within sites (Munkittrick et al., 2002). If an effect was found in one sex of one species, in over half the cases (91 of 167) the same effect was found in the same sex of a second species at that site.

Figure 5-1: Comparison of female (black box) and male responses (hatched box) for condition factor (A) and relative changes in liver weight (B) and gonad weight (C)

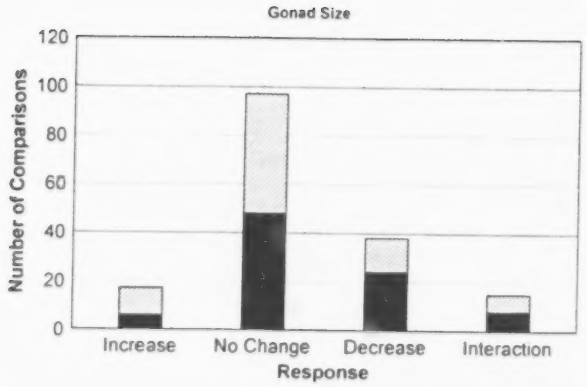


(A)

(B)



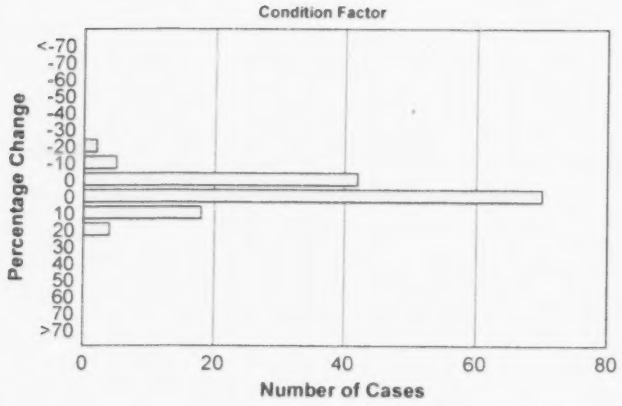
(C)



Percent changes in condition, gonad weight, and liver weight (excluding data with a significant interaction) were plotted to show the magnitudes of the effects within exposure areas (Fig. 5-2). These effect sizes are expressed as exposure versus reference differences as a percentage of the reference mean (as opposed to the standardized effect sizes used for the meta-analyses; see Chapters 3 and 4 for further discussion). For condition, the percent change ranged from a decrease of 27.4% to an increase of 24.5%. For gonad weight, the percent change ranged from a decrease of 90.3% to an increase of 286%. The percent change in liver weight ranged from a decrease of 87.7% to an increase of 73.4%. Sites with the largest increases and decreases in female gonad weight (including corresponding changes in female liver weight and condition factor and male gonad weight) can be found in Table 5-5, indicating further the magnitude of the effects that were found in Cycle 2.

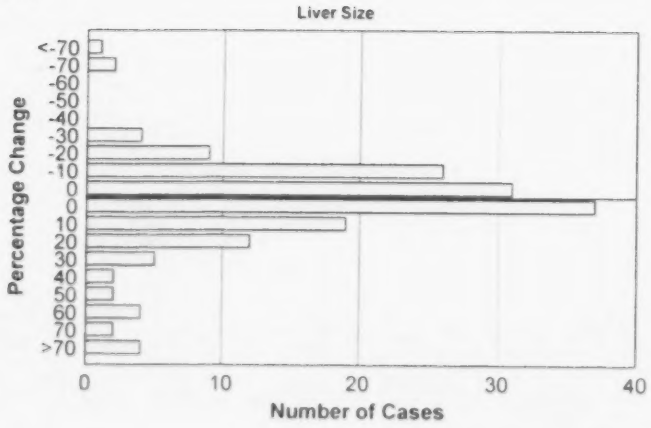
Table 5-6 shows the percent changes in condition, gonad weight, and liver weight that were found in Cycle 2 and the corresponding percentiles that the changes fall into. For example, a 50% decrease in gonad weight falls within the lower 5th percentile.

Figure 5-2: Percent changes in condition factor (A) and relative changes in liver weight (B) and gonad weight (C). Data sets showing interactions were excluded from this analysis.



(A)

(B)



(C)

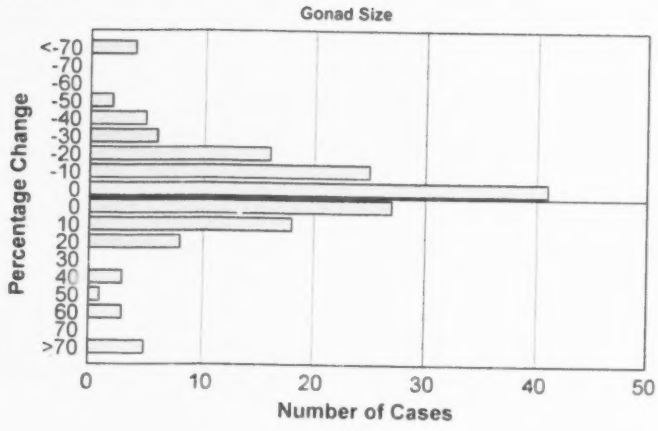

Gonad Size
Percentage Change
<-70
-70
-60
-50
-40
-30
-20
-10
0
0
10
20
30
40
50
60
70
>70
0
10
20
30
40
50
Number of Cases

Table 5-1: Sites with the largest percentage of (A) decreases and (B) increases in gonad weight in female finfish sampled during Cycle 2 (percent difference is given as exposure minus reference mean, expressed as a percentage of the reference mean). For comparison, corresponding changes in male gonad weight, female condition factor, and female liver weight are also given.

A. Decreases in gonad weight

Site	Species	Common Name	Gonad Weight Female	Liver Weight Female	Condition Factor Female	Gonad Weight Male
1.	*N. hudsonius*	Spottail shiner	-90.3	NS	16.1	-83.4
2.	*E. nigrum*	Johnny darter	-47.6	-16.6	-107 (INT)	NS
3.	*C. commersoni*	White sucker	-45.9	-30.3	-5.16	-26.7
4.	*C. catostomus*	Longnose sucker	-45.9	21.3	NS	NS
5.	*P. americanus*	Winter flounder	-33.0	17.2	4.95	NS
6.	*N. volucellus*	Mimic shiner	-32.9	31.2	-7.96	NR
7.	*C. catostomus*	Longnose sucker	-28.9	NS	20.6 (INT)	-25.0
8.	*C. commersoni*	White sucker	-27.6	NS	4.23	59.2 (INT)
9.	*P. flavescens*	Yellow perch	-26.0	-15.1	-10.1	NS
10.	*C. catostomus*	Longnose sucker	-24.0	NS	6.91	NS
11.	*C. catostomus*	Longnose sucker	-23.4	NS	NS	-14.1
12.	*H. alosoides*	Goldeye	-22.6	NS	NS	NC
13.	*C. commersoni*	White sucker	-22.0	41.2 (INT)	NS	NS
14.	*F. heteroclitus*	Mummichog	-19.1	25.3	-6.03	NS
15.	*C. commersoni*	White sucker	-18.3	NS	-28.5 (INT)	-59.1 (INT)
16.	*C. commersoni*	White sucker	-18.0	NS	48.1 (INT)	NS
17.	*C. commersoni*	White sucker	-18.0	NS	6.91	-15.9
18.	*M. dolomieu*	Smallmouth bass	-17.6	NS	3.99	NS
19.	*C. catostomus*	Longnose sucker	-15.9	62.9	-18.3 (INT)	-24.3
20.	*C. commersoni*	White sucker	-14.7	NS	NS	NS
21.	*C. commersoni*	White sucker	-12.7	NS	NS	-11.1
22.	*C. commersoni*	White sucker	-11.3	NS	NS	-49.5 (INT)

B. Increases in gonad weight

Site	Species	Common Name	Gonad Weight Female	Liver Weight Female	Condition Factor Female	Gonad Weight Male
23.	*C. catostomus*	Longnose sucker	60.3	139	18.9	4.5
24.	*R. cataractae*	Longnose dace	25.1	NS	NS	21.3
25.	*C. ricei*	Spoonhead sculpin	57.0	62.2	30.9 (INT)	10.9
26.	*C. catostomus*	Longnose sucker	67.5	69.8	18.5 (INT)	1.63

Site	Species	Common Name	Gonad Weight Female	Liver Weight Female	Condition Factor Female	Gonad Weight Male
27.	*C. commersoni*	White sucker	43.5	17.5	NS	-31.2 (INT)

Note: All percentages are calculated from detransformed adjusted means, except where there is a significant interaction (INT), in which case the percentage is calculated from log-transformed slopes. NS = not significantly different; NR = not reported.

Table 5-2: Percent change in condition, gonad, and liver weight and the corresponding percentile

Percentile	Percent Change		
	Condition	Gonad	Liver
99	18.4	97.6	106
95	16.4	43.6	58.3
90	10.6	17.9	32.13
75	6.05	4.93	17.4
50	2.00	-3.99	3.04
25	-1.84	-16.0	-9.74
10	-4.90	-27.7	-16.8
5	-8.04	-45.9	-27.9
1	-16.4	-83.2	-62.5

5.4.2 Interpretation of ANCOVA Results

It should be noted that not all the measured effects (significant exposure versus reference differences) found in Cycle 2 were necessarily mill-related effects. There were some sites where concerns with the study design resulted in responses that were questionable. For example, one study sampled the exposed area one month after sampling the reference area. A second study used a reference area upstream of a dam. Both of these sites had effects in the three key endpoints, in both sexes and both species, however the differences may have been related to problems with the study design rather than the mill effluent (Munkittrick et al., 2002).

Moreover, there is a difference between finding an effect and the effect being ecologically relevant. Munkittrick et al. (1997, 1998) and McMaster et al. (1991) found a number of effects in fish at Jackfish Bay on Lake Superior, including decreased gonad size, altered sexual characteristics, and delayed sexual maturity. Yet the fish community is sustaining itself, indicating that the population can survive despite the presence of numerous impacts. It was suggested that the fish in the area seek refuge in adjacent bays for reproduction. Therefore, recruitment of juveniles into the exposure area from the adjacent bays may support the population, which illustrates the potential complexity of reproductive effects when fish are exposed to effluent.

An effect in the EEM program is defined as a statistically significant difference in measurements between an exposure area and a reference area. As such, an increase or a decrease in the selected effect endpoints would constitute an effect. Nationally, the overall median response was an increase for condition and liver weight and a decrease for gonad weight in fish exposed to effluent. An increase in condition factor was found more than twice as often as a decrease. For liver weight, in 40% of the cases, an increase was found rather than a decrease. A decrease in gonad weight was found twice as often as an increase. In approximately 10% of the cases for condition factor, liver weight, and gonad weight, there were significant interactions (non-parallel exposure versus reference area slopes) (Fig. 5-3).

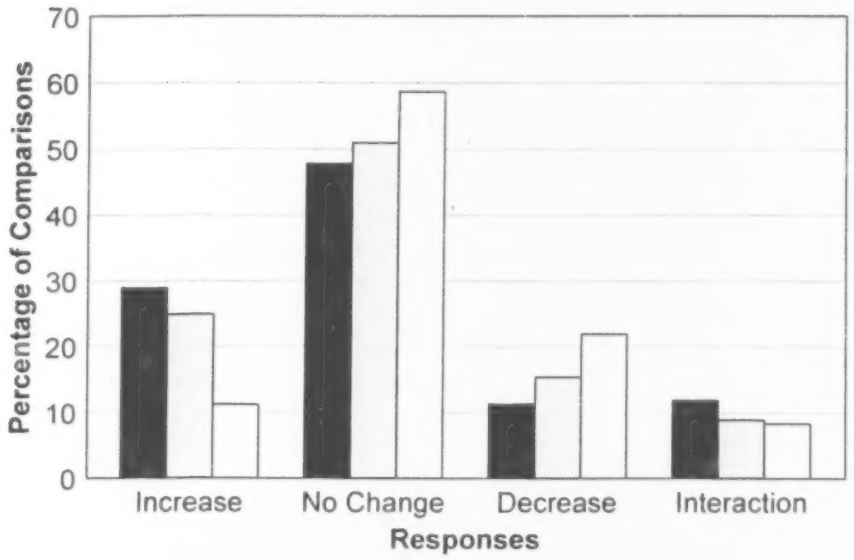


Figure 5-1: Cycle 2 fish responses for condition (black, n = 140), liver weight (shaded, n = 155), and gonad weight (clear boxes, n = 154). Values are given as the relative difference between the exposure versus reference area means, expressed as a percentage of the reference area mean.

Various monitoring frameworks have been developed to aid in identifying responses within a fish population, and this information has been used to focus subsequent monitoring based on the mechanism of the impact (Colby and Nepszy, 1981; Colby, 1984; Munkittrick and Dixon, 1989a,b; Gibbons and Munkittrick, 1994). Gibbons and Munkittrick (1994) suggested that response patterns can form the basis of interpretation of the results of preliminary sampling. Inspection of the overall results can help with identification of a response pattern, the probable mode of action on the fish population, and the strategy for follow-up studies. The authors acknowledged that identification of response patterns may not provide information on the direction in which or speed with which patterns are changing.

Gibbons and Munkittrick (1994) developed a framework that uses response patterns from information on age structure (mean age or age distribution), energy expenditure (growth rate,

gonad weight, fecundity, or age at maturity), and energy storage (condition or liver weight). Resulting patterns have been used to describe changes, including differences in food availability, adult mortality, recruitment failure, and metabolic disruption. In all cases, careful selection of the exposed and reference areas is needed in order to obtain meaningful results from the framework.

Fish measurements can be grouped based on energy use (growth rate, gonad weight, fecundity, age at maturity), energy storage (condition, liver weight), and age (mean age or age distribution) (Munkittrick et al., 2002). Response characteristics of fish in an exposure area can be identified by an increase (+), a decrease (-), or no difference (0) in a response endpoint, compared with fish in a reference area. Generalized response patterns can then be identified that may identify causal factors for the effects being detected (Gibbons and Munkittrick, 1994).

For example, a pattern of (000), representing age, energy use, and energy storage, would suggest a fish population unaffected by the effluent (Gibbons and Munkittrick, 1994). A response associated with eutrophication and increased food resources would have a pattern of (-++). The decrease in age structure would be related to the fish population having younger fish because of an increase in reproduction, survival, growth, and condition because of abundant food resources (Munkittrick et al., 2002). The response patterns are a useful tool in determining what subsequent monitoring may be useful, identifying the cause of effects, and determining follow-up action (which falls outside the EEM program).

Based on the ANCOVA results, it appears that, on a national level, the fish downstream of pulp mill effluents are experiencing a combination of metabolic disruption (a change in the ability to process energy; Munkittrick et al., 2000) and nutrient enrichment (an increase in food availability). Although the quantity and quality of Cycle 1 data were limited, the general pattern that was found in Cycle 2 is consistent with that of Cycle 1.

While increased liver size is commonly attributed to increased activity of liver detoxification systems, it is unclear why liver sizes are usually larger downstream of pulp mills. There have been sites where mixed-function oxidase induction has occurred in the absence of changes in liver sizes and sites where liver size increases have occurred without mixed-function oxidase induction (Munkittrick et al., 1994, 2000). These larger livers are commonly higher in lipids or glycogen (Servos et al., 1994), and it may be that the larger livers are a reflection of altered energy storage and not due directly to increased detoxification enzymes.

Compared with the results extracted from the mill reports, fewer interactions were found in the electronic data for Cycle 2 (fewer cases where the exposure versus reference areas had significantly different regression slopes). This was likely partly due to the removal of immature fish and other outliers from the data set. It is important to note that, when an interaction is found, the data are interpretable — the interaction is the effect. If, for example, there is an interaction in gonad weight versus body weight between exposed and reference populations, this suggests that effluent exposure affects energy utilization for gonadal growth differently for fish of different sizes. This is as significant as effluent exposure causing the same effect for fish of all sizes (i.e., slopes parallel). The form of the interaction may also suggest when populations are affected, such as when they first reach maturity (first spawning) or even as they get older (e.g., as might be seen for weight at age ANCOVAs).

When interpreting the meta-analyses in the following section, it is helpful to see the non-standardized effect size medians (the meta-analyses use standardized effect sizes). Figure 5-4 gives these medians for the four ANCOVA endpoints (calculated as the exposure mean minus the reference mean, divided by the reference mean, using detransformed values; calculated with sign, or direction of effect, included for each mill value). Using the median, exposure to effluent was associated with a 2% increase in condition factor, a 5% decrease in gonad weight, a 3% increase in liver weight, and a 5% increase in weight at age.

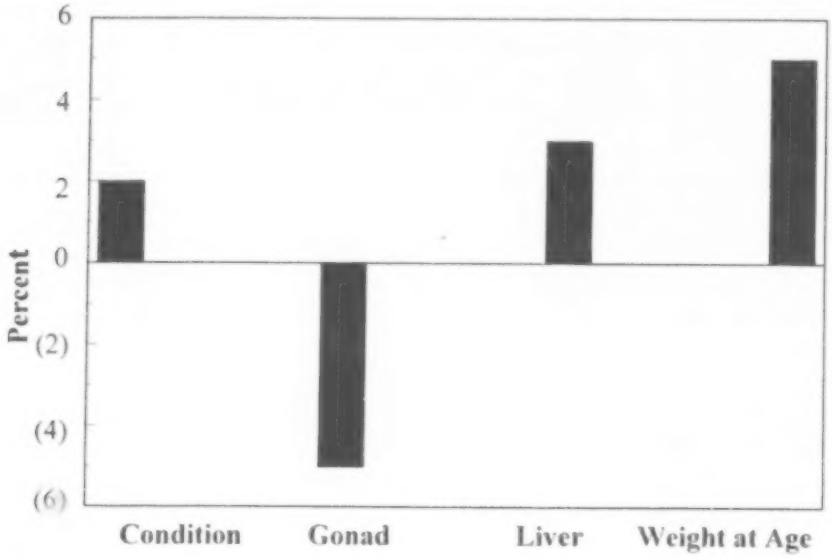


Figure 5-2: Detransformed percent changes in effect size relative to the reference mean

5.5 National Trends — Meta-analyses

5.5.1 Results from Meta-analyses

The meta-analyses were conducted using the corrected electronic data that were submitted by the mills. Only data where slopes were parallel (i.e., no interactions) were used in the meta-analyses. The key fish endpoints that were used in the meta-analyses were gonad weight, condition factor, liver weight, weight at age, and age. The results are presented as Hedges' d standardized effect sizes. These were evaluated relative to potential influencing factors, which, for the fish population survey, were gender, fish species, mill process type, effluent treatment type, and habitat of the receiving area. As for the other sections of this report, this section presents national average patterns; individual mills sometimes have site-specific considerations that may suggest modifications for a particular mill of the interpretations described below.

5.5.1.1 Gender

Male and female fish had similar responses for each of the endpoints, with the exception of gonad weight (Fig. 5-5). Condition, liver weight, and weight at age were all significantly increased in the exposure area for both males and females. Female fish found in the exposure area were significantly older. Male fish in the exposure area were also older, however not significantly. Both males and females had significantly smaller gonads in exposure areas. The magnitude of this decrease was significantly greater for females (n = 66) than for males (n = 60) (P = 0.008). The cause of this gender difference is currently unknown, but could have been related to factors such as less synchronized gonadal growth in males, which might have resulted in a lesser magnitude of decrease relative to background variability.

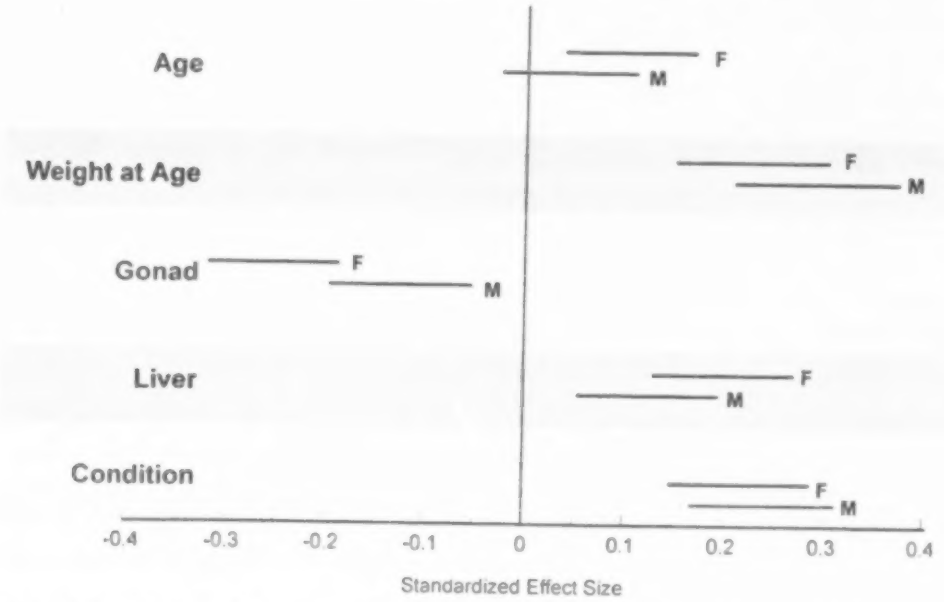


Figure 5-1: Effect of gender on the five key fish endpoints. M = male, F = female. Error bars represent 95% confidence intervals

5.5.1.2 Species

Comparisons were also made among the 19 fish species for which data were available (Table 5-7; 16 freshwater and 3 marine fish species). Not all species had results for all response variables, due to interactions or measurements not being taken.

Table 5-1: List of all species that were included in the meta-analysis

Species	Common Name
Catostomus commersoni	White sucker
Catostomus catostomus	Longnose sucker
Percopsis omiscomaycus	Trout-perch
Perca flavescens	Yellow perch
Etheostoma nigrum	Johnny darter
Ameiurus nebulosus	Brown bullhead
Catostomus macrochelius	Largescale sucker
Etheostoma exile	Iowa darter
Rhinichthys cataractae	Longnose dace
Myoxocephalus scorpius	Shorthorn sculpin
Gasterosteus aculeatus	Threespine stickleback
Etheostoma olmstedi	Tesselated darter
Proposium williamsoni	Mountain whitefish
Cottus ricei	Spoonhead sculpin
Rhinichthys atratulus	Blacknose dace
Micropterus dolomieu	Smallmouth bass
Pleuronectes americanus	Winter flounder
Fundulus heteroclitus	Mummichog
Ambloplites rupestris	Rock bass

Figure 5-6 illustrates the meta-analysis plot of the four most commonly used species (white sucker, longnose sucker, yellow perch, trout-perch) for each of the five key endpoints. For the other 15 fish species that were included in the meta-analysis, data were generally available only for 2–4 comparisons per endpoint (sampled in one or two studies), and it is not possible to draw firm conclusions at this time for these less commonly used species. The data that were available suggested that some of these latter species may have experienced large effects, however, and may be deserving of further study.

Nationally, the effects found in the two most commonly used species (white sucker, n = 39–47; longnose sucker, n = 18–28; n represents the number of comparisons, which was variable between endpoints because of interactions and missing data) were a significant decrease in gonad weight and a significant increase in liver weight, condition factor, and weight at age. The longnose sucker had a significant increase in age in the exposure area, but the average age structure for white suckers was not significantly different in the exposure and reference areas. The effects found in the longnose sucker were consistent with the national grand means for all five endpoints.

Trout-perch (n = 5–8) and yellow perch (n = 5–6) were sampled at much fewer sites than either the white or longnose sucker. Nationally, yellow perch did not follow the same trend as the suckers — the perch sampled in the exposure area had a significantly decreased weight at age, condition factor, and gonad weight. Liver weight was also decreased, although it was not significantly different from the reference area. Age was significantly increased in the exposure

area. Trout-perch also had a unique response pattern. Liver weight was found to be significantly decreased, but that was the only endpoint where there was a significant change compared with reference areas.

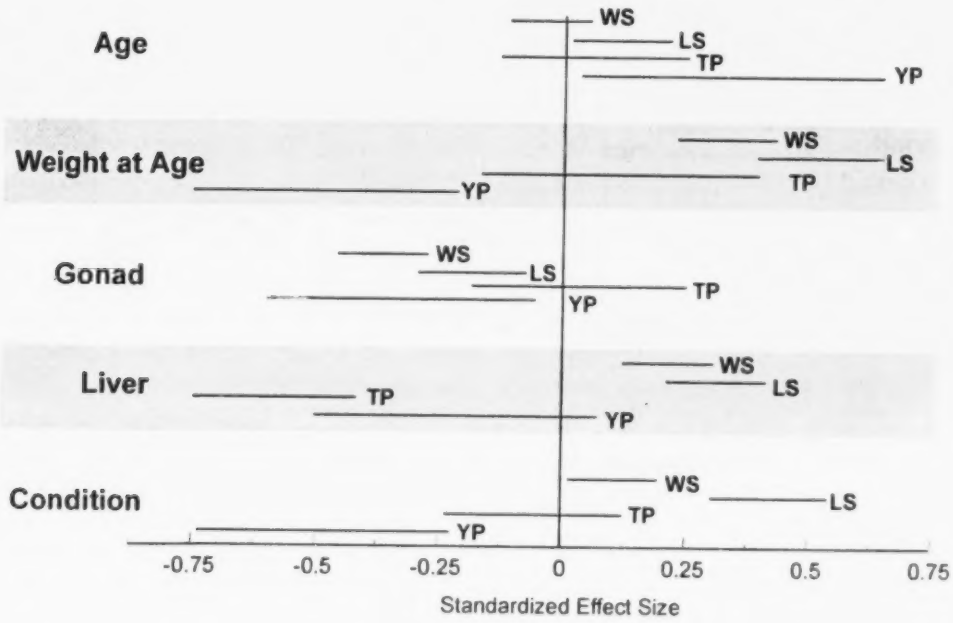


Figure 5-1: Effect of species on the five key fish endpoints. WS = white sucker; LS = longnose sucker; TP = trout-perch; YP = yellow perch. Error bars represent 95% confidence intervals

5.5.1.3 Mill Process Type

A number of mill process types were identified by the mills that submitted electronic EEM data. These were grouped into eight general process types. The most common was the kraft process (n = 37–48), followed by the thermomechanical pulp process (n = 33–38). The meta-analysis summary for these two mill process types is illustrated in Figure 5-7. The other six mill process types were less commonly used, so definitive conclusions are not warranted at this time, although initial data suggest that some of the process types should receive further study. Nevertheless, since the thermomechanical pulp and kraft processes represent the majority of mills (according to the way in which mills were classified for this assessment), the results do give a good indication of the overall national trends.

Mills using a thermomechanical pulp process type had significant differences in all five key endpoints, which were consistent with the national grand mean (decrease in gonad weight and increase in age, weight at age, liver weight, and condition). Kraft mills were almost identical, with the exception of liver weight; although it showed an increase, it was not significantly different from the reference area.

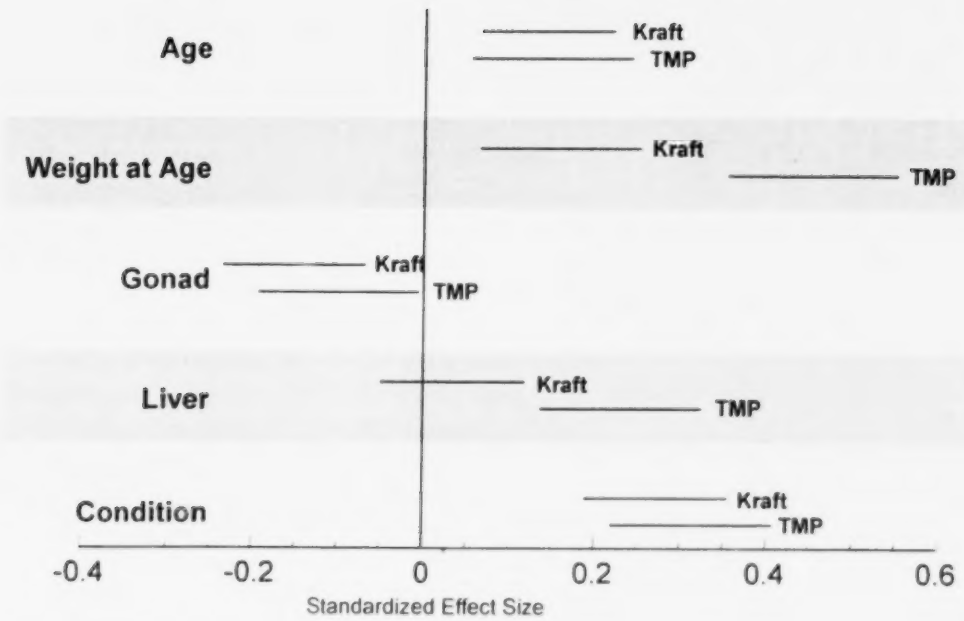


Figure 5-1: Effect of mill process type on the five key fish endpoints. Kraft = kraft mills; TMP = thermomechanical pulp mills. Error bars represent 95% confidence intervals

5.5.1.4 Treatment Type

Four different effluent treatment types were identified by the mills that had submitted electronic data. The greatest number of comparisons was for mills using activated sludge treatment (n = 59–78) and aerated stabilization basin treatment (n = 33–42). The results of the meta-analysis for these two treatment types can be found in Figure 5-8. Only a few mills in Cycle 2 reported results from a fish population survey conducted where primary or sequential batch reactors were used to treat effluent, making it difficult to draw firm conclusions at this time. Again, the initial results suggest that these latter groupings should be studied further.

Both activated sludge and aerated stabilization basin treatment followed the national trend, with a statistically significant increase in weight at age, liver weight, and condition and a statistically significant decrease in gonad weight in fish sampled in the exposure area. Age was increased at sites with activated sludge and aerated stabilization basin treatment, although this difference was not significant for aerated stabilization basin treatment.

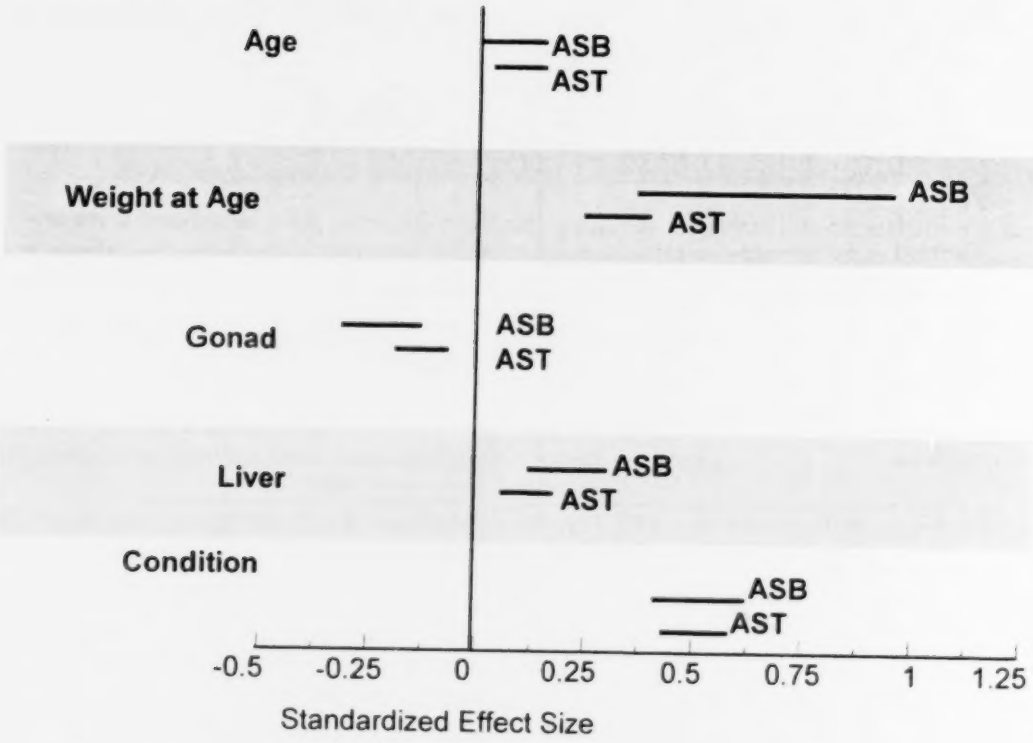


Figure 5-1: Effect of treatment type on the five key fish endpoints. ASB = aerated stabilization basin; AST = activated sludge treatment. Error bars represent 95% confidence intervals

5.5.1.5 Habitat

Habitat type was broken down into three types of receiving areas: riverine, lacustrine, and marine/estuarine. All three habitat receiving areas demonstrated a significant decrease in gonad weight, with fish collected in lakes and marine/estuarine sites exhibiting the greatest decreases (Fig. 5-9).

Condition was significantly increased for all three habitat types. Liver weight was significantly increased for rivers and marine/estuarine sites, with no significant difference in lakes. Weight at age was significantly increased for rivers and lakes, with no significant difference in marine/estuarine sites. Age was significantly increased for rivers and significantly decreased in lakes and marine/estuarine receiving areas.

Rivers had significant changes in all five response variables, while lakes and marine/estuarine habitats had significant changes in four of the five variables.

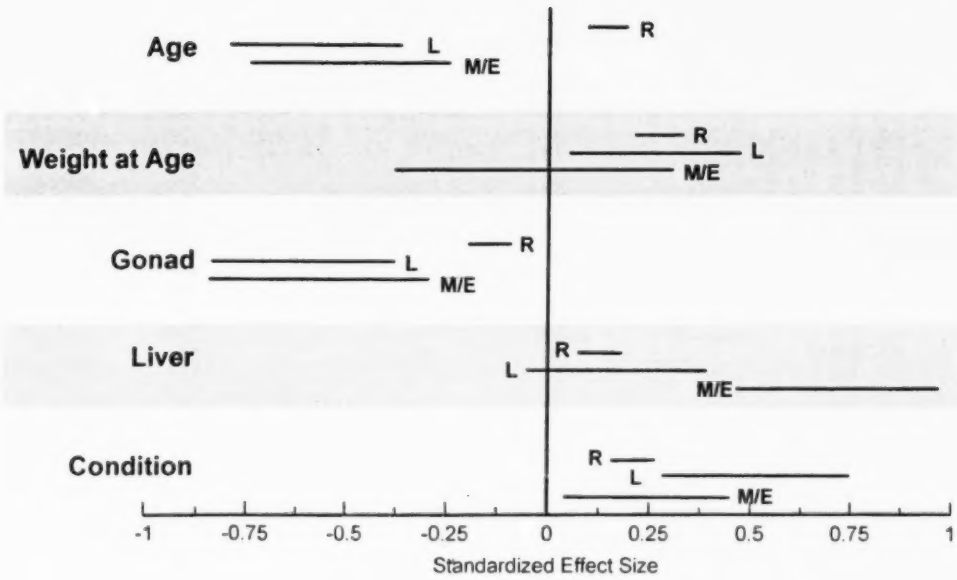


Figure 5-1: Effect of habitat on the five key fish endpoints. R = riverine; L = lacustrine; M/E = marine/estuarine. Error bars represent 95% confidence intervals

5.5.2 Interpretation of Meta-analyses

Despite "noise" in the data due to site-specific differences among mills, certain national patterns were apparent. These average responses found during Cycle 2 will be used for comparisons with subsequent cycles to determine if the magnitude or direction of effects is changing.

The meta-analyses showed that the results for males and females for the five endpoints were generally consistent. Both male and female fish had gonad weights that were significantly reduced relative to reference fish, with female fish gonad weights reduced to a greater degree than those of male fish. On a national level, both liver weight and the condition of male and female fish were increased in exposed populations. Although fish were generally older in exposed populations, only female exposed fish were significantly older than reference females. A number of studies conducted near pulp and paper mills looking at the same type of population endpoints as EEM have found the same degree of consistency in responses between male and female fish. At this point, there is uncertainty concerning the cause of the increased age response in exposure areas, although it could be due to some combination of loss of juveniles, greater survival of older fish, or delayed maturity in female fish.

The two most common species used in EEM studies (white sucker and longnose sucker) had results that were generally consistent with the national trends: decreased gonad weight and increased liver weight, condition factor, age, and weight at age. It is important to point out that there were exceptions to these patterns. For instance, the national results for yellow perch were consistent with what is typically found under conditions of food limitation. Gibbons and

Munkittrick (1994) suggested that a population with lowered food resources tends to reduce its energy expenditure (slower growth and reproduction), without an initial impact on age structure. However, the EEM results for perch showed that fish in the effluent were older, which may indicate increased juvenile mortality or delayed sexual maturity (Gibbons and Munkittrick, 1994).

There are a number of other potential explanations for the different responses between sucker and perch, including site differences in such factors as effluent characteristics and receiving environment characteristics, as well as differences in life history strategies. Sucker species generally eat the same organisms throughout their life cycle, whereas yellow perch usually switch from feeding on zooplankton to piscivory when they reach a suitable size. It is possible that the perch captured in the effluent were changing trophic levels (from benthivore to piscivore) later than the reference fish. Another possibility is that the fish captured in the reference area may not have been comparable for other reasons.

Different species are sometimes expected to respond differently at sites exposed to the same effluents. Gibbons et al. (1998) compared the results of trout-perch and white suckers downstream of a pulp mill in Northern Ontario. At that particular site, older, larger trout-perch were not present, whereas juvenile mortality appeared to be more prevalent for white sucker populations. The interpretation in this case was that both species may have been responding to predation by walleye, and that the mill appeared not to be the dominant influence on fish population responses downstream of the mill. This was in contrast to earlier studies at that site (before modernization and secondary waste treatment) that suggested that fish did show reproductive impairment associated with effluent exposure (Munkittrick et al., 1994, 2000).

Riverine habitat was by far the most common habitat sampled during Cycle 2 EEM studies, and, on average, fish in rivers exposed to effluent showed increases in condition factor, liver weight, growth rate, and age and decreases in gonad weight. As expected, given the predominance of river habitats sampled, this pattern reflected the overall national trend. Marine/estuarine and lacustrine habitats were sampled much less frequently, and more detailed interpretation of those results awaits the availability of further data.

On a national basis, effluent-exposed fish were found to be slightly older, with faster growth, larger livers, increased condition, and reduced gonad weights. These are the same trends that were originally seen in detailed research studies on individual mills more than 10 years ago (McMaster et al., 1991; Munkittrick et al., 1991, 1994). This national average response pattern was obtained even after including a variety of sites where overall eutrophication was the dominant response pattern (including increased gonad weight), as well as some sites where the site-specific response was related to food limitation (including decreased condition and liver weight).

The predominant response pattern seen (decrease in gonad weight and increases in liver weight, condition, and weight at age) is believed to be indicative of some form of metabolic disruption (in combination with a nutrient enrichment effect). This interpretation was supported by the tabulations of ANCOVA results presented in the previous section, and this response pattern has also been found in other studies conducted near pulp mills (Munkittrick and Dixon, 1989b;

McMaster et al., 1991; Munkittrick et al., 1991, 1994, 2000). Although the mechanism by which these responses are occurring (as well as how important these responses are ecologically) is unclear at this time, these consistent patterns of effects on a national scale may be related to some form of disruption of endocrine functioning. Further research to understand this phenomenon is required, in combination with EEM studies to confirm the type, spatial extent, and ecological importance of these effects.

5.6 Evaluation of Selected Fish population survey Program Elements

5.6.1 Effect Size for Calculating Required Sampling Effort

A review of the variability observed in the fish endpoints following Cycle 1 showed large differences between studies, even those at closely related sites (Environment Canada, 1997c). A target effect size was established after Cycle 1 to assist in the evaluation of the Cycle 1 results. Although an expert working group looking at the fish population survey did not come to a consensus on the target effect size, the government members of the fish population survey expert working group felt a range of 20–30% for the target effect size was appropriate, with the lower limits set based on two considerations. First, as the target effect size is reduced, the sample size increases, which may create effects based on sampling mortality that exceed effects from the effluent. Second, if the target effect size is too small, there is a risk of detecting a natural difference and attributing it to the effluent (Type I error).

Using the electronic data, the percentage of fish comparisons where the measured effect size exceeded 2SD, 25%, and 10% was examined (Fig. 5-10). During this exercise, it was noted that not all parameters measured for the fish population survey are equally variable and that reproductive endpoints have been found to be the most variable on a relative scale (Environment Canada, 1997c). Effect size information was not collected with the expectation that an effect of a particular magnitude would necessarily be considered of ecological, economic, or social significance, but rather to determine the effect size that would be used *a priori*, in order to calculate fish sampling effort. The goal in setting a predetermined effect size is to determine (using power analysis calculations) the sampling effort required to detect the worst-case scenarios. For liver weight, gonad weight, and weight at age, approximately 20% of the mills exceeded a predetermined effect size of 25%. Therefore, the Cycle 2 data show that an effect size of 20–30% for gonad weight (which is what was recommended following the Cycle 1 review), as well as liver weight and weight at age, would likely be appropriate for use in a power analysis for determining sample sizes *a priori* in EEM. For condition factor, approximately 20% of the mills exceeded a predetermined effect size of 10%, indicating that this lower effect size would be more appropriate for calculating required sampling effort for condition. A more detailed discussion on *a priori* effect size determination and power analysis is found in the previous chapter of this report.

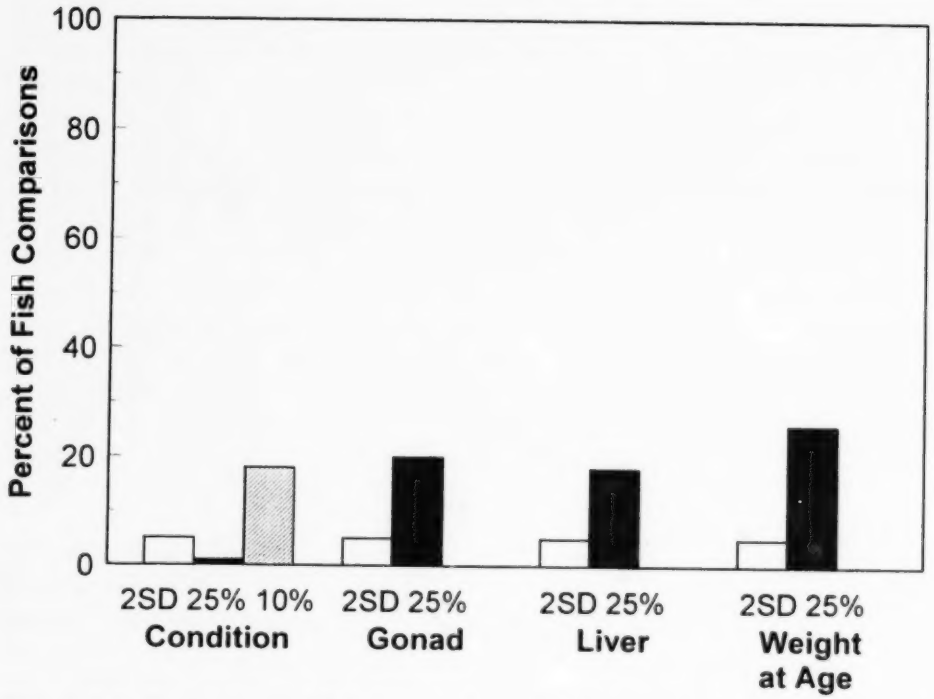


Figure 5-1: Percentage of fish comparisons in Cycle 2 where the measured effect size exceeded two standard deviations (2SD), 25%, or 10%

5.6.2 Evaluation of Numbers of Significant Differences for Each Endpoint

The number of statistically significant endpoints observed at each mill was examined, in part to address the issue of finding significant differences in the individual mill comparisons due to random chance, given the moderately high number of comparisons performed at each mill for the fish population survey (Fig. 5-11). Figure 5-11 shows that less than 10% of mills had no significant differences in any of the three key endpoints examined (condition, gonad weight, liver weight). More than 40% of mills had at least one statistical difference (for up to two sexes and species per mill) for all three endpoints. Therefore, for Cycle 2, because there were very few mills where there were effects on only one endpoint, the results suggest that the overall pattern of effects that was found was not related to random chance, but was more likely a response to effluent exposure. Finally, Figure 5-11 also shows the good agreement between the results of the analyses of the electronically submitted data and the analyses provided by the interpretative reports.

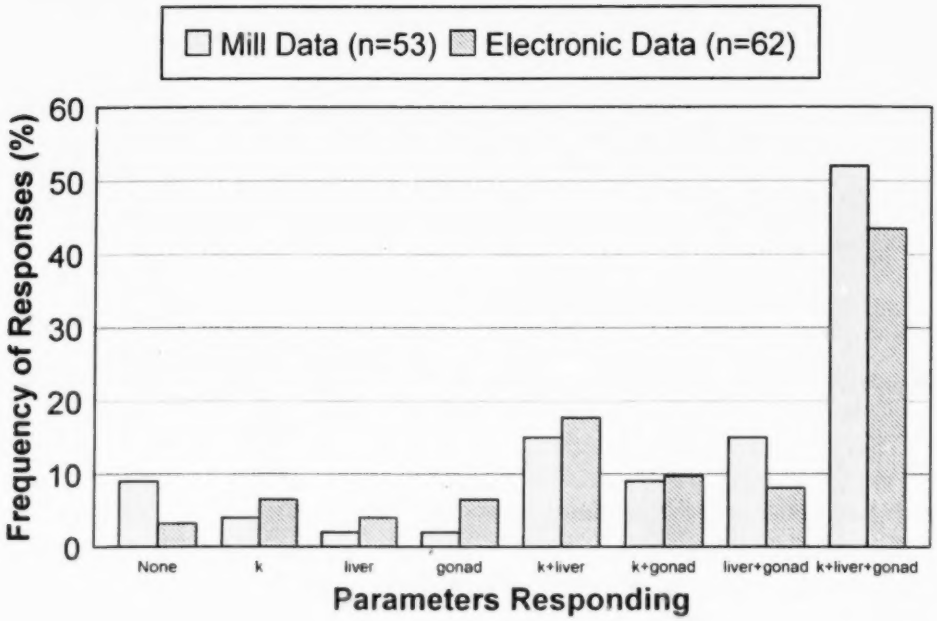


Figure 5-1: Types of statistically significant endpoints observed at each mill site using data extracted from the mill interpretative reports (filled boxes, n = 53) and the electronic data (hatched boxes, n = 62). Results are presented as the percentage of times that mills found specific combinations of these responses — condition factor (k), gonad weight, and liver weight

5.7 Conclusions and Recommendations

The second cycle of EEM proved to be much more successful than Cycle 1 for the fish population survey. In Cycle 2, the majority of mills that attempted a fish population survey were able to capture sufficient numbers of fish of both sexes of at least one fish species, which facilitated obtaining interpretable data that can be used for future comparisons. Part of the increased sampling success may be related to more mills using smaller-bodied fish species (such as minnows, darters, and sculpins), which are generally easier to capture. The smaller fish species are generally less mobile than larger fish species, so there is a greater likelihood of exposure to mill effluent.

There were very few sites in Cycle 2 showing "no effects" in fish. At first look, this information suggests that effects may be increasing. However, the fact that more sites were successful in capturing fish in Cycle 2 complicates matters. Since Cycle 1, fish may be moving back into areas where they were previously absent (e.g., the fish may have previously avoided the areas of effluent exposure). Moreover, by the time of Cycle 2 sampling, secondary effluent treatment had been installed at almost all mills in Canada. The improvement in effluent quality (which is clearly demonstrated by the results of the mill sublethal toxicity tests; Scroggins et al., 2002) may be allowing the fish to survive in areas where they previously could not.

The effects that were found in fish in Cycle 2 were generally consistent between sexes as well as between fish species at the individual sites. In over half of the ANCOVA comparisons, if effects were found in one sex or fish species, the same effects, in the same direction, were found in the other sex or fish species. The consistency of results between sex and fish species was also seen in the meta-analyses and subsequent multivariate ordinations. For example, for many of the mills, the points for males and females fell within close proximity of one another (Chapter 6).

The dominant pattern found in Cycle 2 with both the ANCOVA summaries and the meta-analyses was a decrease in gonad weight and an increase in liver weight, condition factor, age, and weight at age for fish in exposure areas, as compared with reference area fish. This pattern is consistent with a response pattern characterized by a combination of metabolic disruption (reallocation of energy use) and nutrient enrichment (increase in energy storage). Some mills also had a pattern consistent with food limitation (decreases in the key endpoints) and chemical toxicity (increase in liver weight, decrease in condition and gonad weight). See Chapter 6 for further discussion of these response patterns.

It should be noted that not all the patterns that were found were necessarily mill-related effects. For instance, some of the individual mills that found food limitation effects also had problems with their study designs, such as the timing of sampling between the exposure and reference areas. For example, in one situation, the fish were collected at the exposure and reference areas one month apart. The data showed increases in condition factor, liver weight, and gonad weight for both species. In another case, fish were collected from a reference area upstream of a dam. The exposure area fish exhibited poor condition and growth as well as decreased liver and gonad weights. Both of these sites exhibited "effects," but the results were likely associated with the study design.

The summaries provided in this report are a good tool for mills to use to understand where their individual results fit into the national picture, in terms of magnitude of effects as well as patterns of effects. These summaries will be useful for subsequent EEM cycles to monitor trends and patterns, as well as the magnitude of effects, at a national level.

The data interpretation that was reported in the mill interpretative reports was consistent with what was found when we analyzed the electronic data. A problem with trying to use the mill interpretative reports for the summary statistics was the inadequate reporting of results, leading to difficulties with data extraction from the reports (e.g., no summary tables, scatterplots not included). Electronic data quality was also a concern — over 50% of the data sets that were submitted electronically contained outliers, including immatures, or other problems. These problem data were often not used in the interpretative report analyses, but this was usually not apparent from what was submitted electronically, as described below. The quality assurance/quality control of the electronic data for Cycle 2 was quite labour intensive. However, it did improve the data quality for the analysis; for example, there were fewer interactions in the data, as well as fewer extreme results due to the removal of immature fish and outliers.

Using the Cycle 2 electronic data, a recommendation for setting effect size, *a priori*, in order to determine sampling effort was developed. For the fish population survey, the Cycle 1 recommendation of a ±25% effect size remains an appropriate target effect size for gonad and

liver weight, as well as weight at age, based on the Cycle 2 data. For condition, the target effect size should be closer to 10%, as this endpoint is less variable and does not respond as dramatically as the other endpoints to effluent exposure.

The quality assurance/quality control of the electronic data was a major undertaking. In many cases, it appears that data excluded in the analyses for the interpretative reports were included in the electronic data. It is recommended that only data used for analyses be included in the electronic data submission. Furthermore, it is recommended that the mills present summary tables in their interpretative reports. The summary statistics for the fish include the arithmetic mean, median, standard deviation, standard error, and minimum and maximum values. Scatterplots in the interpretative reports of the fish data plots would also be useful in identifying outliers and other unusual data. For example, when reproductive endpoints are analyzed for fish, the plots aid in identifying potential "immature" fish, which could affect the results.

When an interaction in the data is found, these data should not be completely discarded. Interactions are interpretable and can indicate an effect. For example, an interaction in condition factor means the length–weight relationship at the exposure area differed from that at the reference area. These interactions should be reported.

6. Overall National Response Patterns for Benthic Invertebrate Community and Fish population surveys

The national response patterns for both the benthic invertebrate survey and the fish population survey are summarized in the next series of figures. These summary figures were produced by plotting the standardized effect sizes (Hedges' d) generated by the meta-analyses for the Control/Impact mills for both benthos and fish. This made it possible to integrate the findings for these two core elements of the EEM program (Gradient designs were not used for the fish population survey and, thus, are not included in the following analyses).

For all but one of the following summary figures (Fig. 6-1 is the exception), information from three or more dimensions (three or more endpoints) was collapsed into two dimensions to aid in interpretation. This was done using multidimensional scaling ordination to analyze multivariate data matrices of mills (objects) versus standardized effect sizes for each endpoint (attributes). The multidimensional scaling analyses were used to find the "best fit" of mills into a two-dimensional ordination space (Faith et al., 1987; Belbin, 1992; Lowell and Culp, 2002). The mills (points) that are closer together in each of the following figures were more similar in terms of measured endpoint effect sizes than mills that are farther apart. The patterns in mill distribution in the following plots reflect different overall response patterns, as discussed below.

Principal axis correlation vectors (Belbin, 1992) are also plotted on the following multidimensional scaling figures. These were calculated using a multiple linear regression technique that determined the linear relationship between each of the standardized endpoints and the ordination space. Each vector gives the direction of best fit within the ordination space for a particular endpoint. This was an improvement over correlating the endpoints to the ordination axes, because it provided the direction of best fit and was not dependent on the positioning of the axes. The correlation vectors could then be used to interpret patterns in the data, as illustrated in the discussion for Figure 6-2. Furthermore, the vectors help to show correlations (or lack of) among the core endpoints. Because standardized endpoints were used in the analyses, it was possible to look not only at correlations among endpoints within the benthic invertebrate survey and within the fish population survey, but also at correlations between the two surveys.

Using the endpoint vectors (and the main axes in Fig. 6-1), it was possible to divide the two-dimensional space of each figure into sectors representing the major categories of effects that were observed (eutrophication, metabolic disruption, etc.). Some distortion is inevitable when collapsing multidimensional data into two dimensions, and distortion may increase as the number of dimensions (endpoints) to be summarized is increased. This distortion is measured as "stress," and the goal is to maintain stress levels below 0.2 to facilitate reliable interpretation of patterns in a two-dimensional multidimensional scaling plot. In the following figures, the stress level for each of the ordinations was within acceptable limits (<0.2), indicating that distortion was fairly minimal. The result was that, on average, the majority of the individual mill points were in or near the effect sector that would be expected, based on the results for each individual mill.

6.1 *Benthic Invertebrate Community Survey*

The overall national response pattern for effluent-exposed benthic invertebrates is captured in Figure 6-1. This figure focuses on the two endpoints used most heavily for interpreting patterns: abundance and taxon richness. The standardized effect sizes generated by the meta-analyses for these two endpoints are plotted on the x-axis and y-axis, respectively. Each point represents a different mill, with different symbols used for the four major habitat categories.

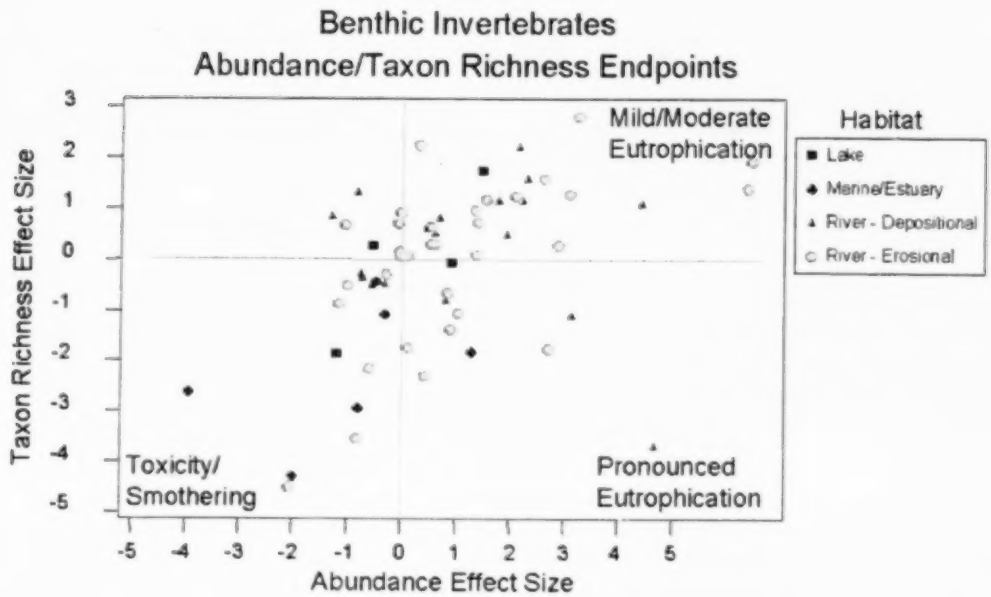


Figure 6-1: Overall national response pattern for benthic invertebrates: abundance and taxon richness endpoints

Exposure areas for mills in the upper right quadrant exhibited increases in both abundance and taxon richness (Fig. 6-1), a characteristic mild to moderate eutrophication response (see also Fig. 4-3 and discussion in Section 4.4.1). Mill exposure areas in the lower right quadrant showed increases in abundance but decreases in taxon richness, typical of more pronounced eutrophication. Mill exposure areas in the lower left quadrant displayed decreases in both endpoints, characteristic of a toxicity/smothering-type response. As illustrated in Figure 4-3, it is more unusual to see stressed communities showing decreases in abundance but increases in number of taxa. Usually, when impacts are great enough to cause decreases in abundance, taxon richness is also decreased. Thus, as expected, there were few mill exposure areas falling in the upper left quadrant. In all four quadrants, points falling farther from the centre represent effects of greater magnitude, while those closer to the centre signify lesser magnitude or no effects.

Hence, Figure 6-1 summarizes the overall patterns discussed in more detail in Section 4.4. Whereas some mills representative of each of the habitat types can be found in each of the quadrants in most cases, certain patterns were apparent. The majority of river mills fell in the right, eutrophication half of the figure, particularly the upper right quadrant, representing mild/moderate eutrophication in exposure areas. In contrast, marine/estuarine mills were clustered more strongly into the toxicity/smothering quadrant, a tendency that would be even more pronounced when including Gradient mill results. Lake mills, in turn, were more intermediate in position, compared with the riverine and marine/estuarine mills (but see also discussion in Section 4.4.2 of pronounced changes in community composition in lake exposure areas).

To determine the relative degree of correlation among the four core benthic endpoints, a multidimensional scaling ordination was performed to reduce the findings for the four endpoints to two dimensions, as discussed above (Fig. 6-2). This helped to ascertain whether, for example, mill exposure areas showing pronounced increases in one endpoint also regularly showed similar increases in another endpoint. The relative degree of correlation among the endpoints is illustrated by the positioning of the principal axis correlation vectors described above.

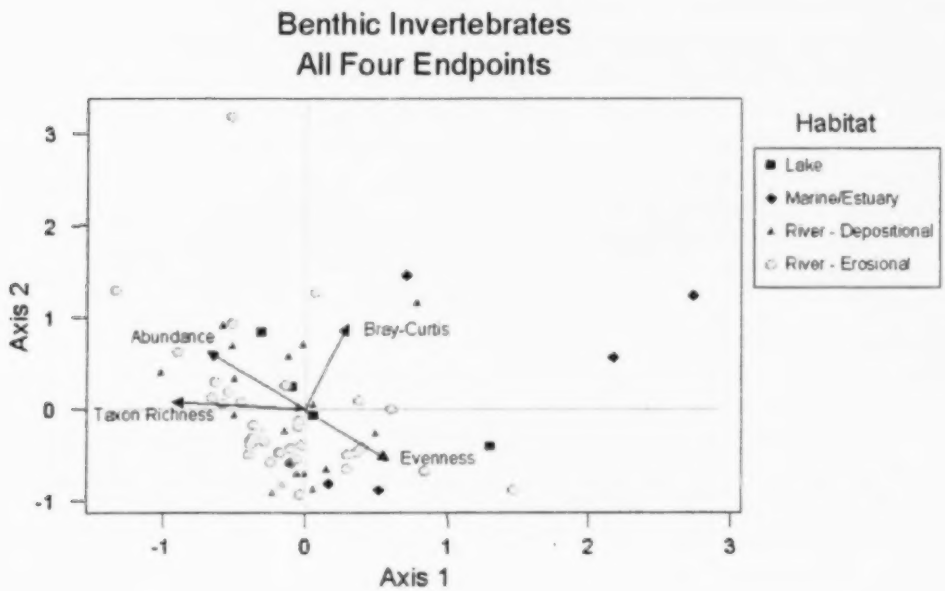


Figure 6-2: Overall national response pattern for benthic invertebrates: all four endpoints

To give an example, the abundance vector in Figure 6-2 gives the direction of best fit for the abundance endpoint effect sizes, as averaged across all the mills. Mills showing pronounced increases in abundance in exposure areas (relative to reference areas) tended towards the upper left of the ordination space, while those showing pronounced decreases in abundance tended

towards the lower right. The taxon richness and evenness vectors can be interpreted in a similar fashion (in terms of positioning of mills with increases versus decreases). The Bray-Curtis endpoint, in turn, was usually positive. Hence, mills showing more pronounced changes in community composition in exposure areas tended towards the upper right of the ordination space, whereas those showing less pronounced changes in community composition tended towards the lower left. So, for example, lake mills showed more pronounced changes in community composition, on average, as also noted in Section 4.4.2. Of note, the evenness vector indicated an average increase in this endpoint for mills showing decreases in taxon richness. As discussed in Section 4.4.2, this may have been related to the loss of more rare taxa for these mills, leaving behind a fewer number of more evenly distributed, pollution-tolerant taxa. Marine/estuarine mills typically showed this kind of response (Figs. 4-12, 4-13, and 6-2).

If one or more of these endpoints were highly correlated for most mills, the principal axis vectors would approximately overlap one another. This might indicate that the two (or more) endpoints were redundant in terms of the mills identified as having pronounced effects in a certain direction. It would not, however, necessarily mean that they were measuring the same effect. For example, two endpoints might be highly correlated if the effect being measured by one endpoint was always associated with a different effect being measured by the other endpoint.

The fact that the four vectors did not overlap indicates that the four endpoints were not redundant. Each endpoint was originally selected to measure a different aspect of the way in which benthic communities might respond to pulp mill effluent, and it appears that this was achieved. There was some correlation among the different endpoints, but this was only partial. Furthermore, the lack of overlap also indicates that the type of community response measured by one endpoint was not always associated with a similar direction of response for any of the other endpoints. This suggests that each of the measures had value for describing a different aspect of the nature of effluent effects. Deletion of any of the endpoints would run the risk of missing an important component of effluent effects, and such deletion is not recommended at this time.

6.2 Fish population survey

The overall national response pattern for the fish population survey is shown in Figure 6-3. Three endpoints (gonad weight, liver weight, condition factor) were particularly useful for interpreting fish responses to effluent exposure, as illustrated in this figure. Multidimensional scaling ordination was used to reduce the meta-analysis-generated effect size findings for these three endpoints to two dimensions. Each point represents a different exposure versus reference comparison. Because the fish population survey incorporated as many as two species and two sexes per mill, this figure (and the following three figures) includes up to four points per mill. Only those mills reporting data for all of the above three endpoints are included in Figure 6-3. Thus, it was necessary to exclude some mills from the analyses. This was also true for the following three figures, for those mills where data were not available for all of the endpoints included in the analyses for each figure. As for Figure 6-2, the principal axis correlation vectors are included to show the direction of best fit for each of the endpoints. The positioning of these vectors helps to identify sectors of the ordination space corresponding to the different main response patterns exhibited by effluent-exposed fish. These response patterns are described more

fully in Chapter 5. As in the previous two figures, points falling farther from the centre signify effects of greater magnitude than those closer to the centre.

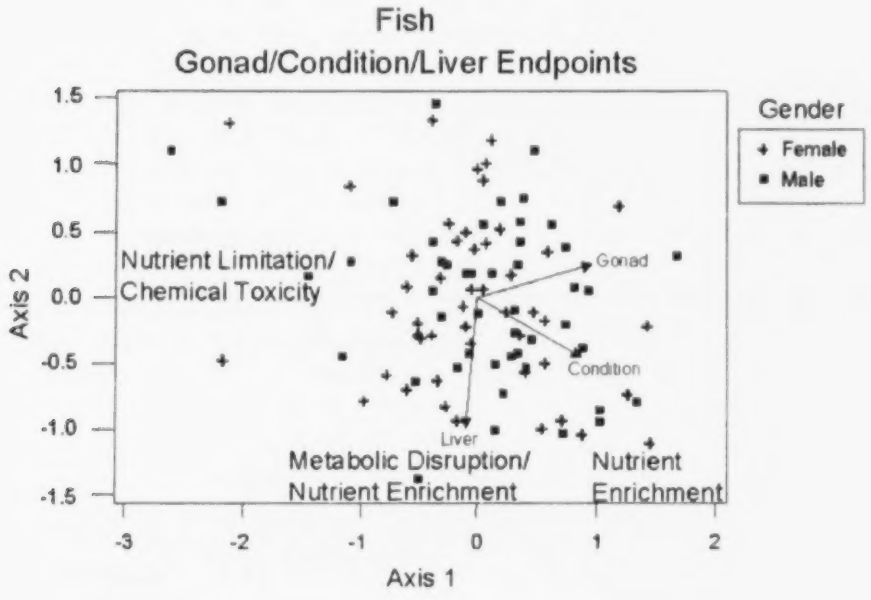


Figure 6-1: Overall national response pattern for fish: gonad weight, condition factor, and liver weight endpoints

Fish exposure versus reference comparisons falling in the lower sector of the ordination space showed the most pronounced combined metabolic disruption/nutrient enrichment responses (decreased gonad weight, increased condition factor and liver weight) (Fig. 6-3). Comparisons in the lower right sector demonstrated more of an overall nutrient enrichment response pattern (increases in all three endpoints). Comparisons in the upper left sector showed a combined nutrient limitation/chemical toxicity response pattern. Nutrient limitation is used here in a more general sense to include some combination of limited food, appetite suppression, and/or internal alteration of food absorption (leading to decreases in all three endpoints). This may include some aspect of chemical toxicity, particularly as it grades into a response pattern of decreased gonad weight and condition factor coupled with increased liver weight. A possible fourth response pattern that became apparent during the analyses was one of increased gonad weight, intermediate condition factor, and decreased liver weight (upper right sector of Fig. 6-3). This response may have been due to study design problems (e.g., sampling the reference and exposure areas at different times) or, alternatively, to effects of the effluent on the timing of reproduction (perhaps involving trade-offs between liver and gonad weights). More focused study will be required to further understand this latter pattern. Also of note, when viewing all the comparisons nationally, both males and females exhibited each of the three (or four) response patterns.

As outlined in Section 6.1 for the benthic invertebrate community survey, an overall picture of the degree of correlation among the five core fish endpoints can be seen by reducing the meta-analysis effect sizes for the five endpoints to two dimensions via multidimensional scaling ordination (Fig. 6-4). Again, this helps to show whether large increases in mill exposure areas for one endpoint were regularly associated with similar increases in another endpoint. If all the endpoints were highly correlated, the five principal axis vectors would roughly overlay one another. This, however, was not the case. Even at the coarser scale of Figure 6-4, the only two endpoints that were fairly highly correlated were those for condition factor and liver weight. Moreover, Figure 6-3 shows that, at a finer scale of resolution, even these latter two endpoints were responding differently at different mills. Thus, Figures 6-3 and 6-4, taken together, illustrate a lack of redundancy among the five core fish endpoints. The correlations that did exist among the different endpoints were only partial. Thus, each endpoint appeared to be identifying a different aspect of effluent effects on fish. As discussed in Section 6.1, this demonstrates the value of measuring and reporting all five of these endpoints.

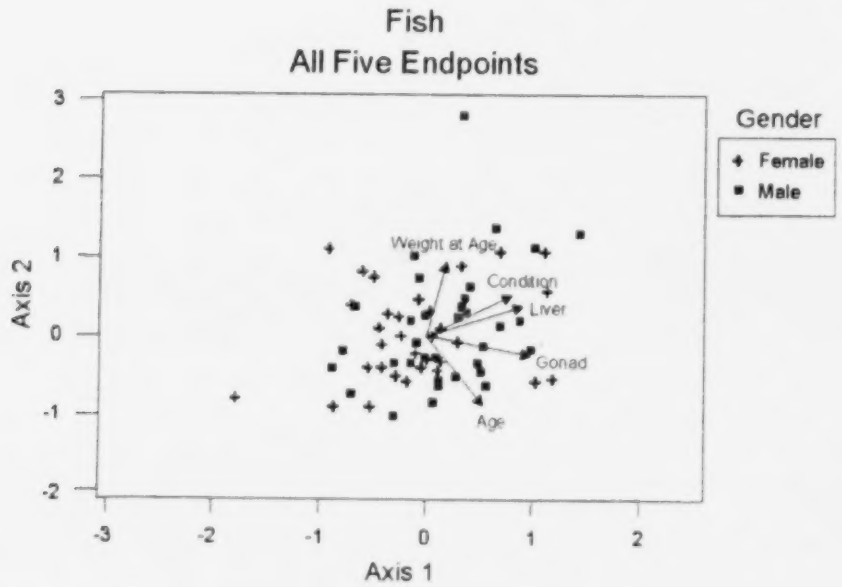


Figure 6-2: Overall national response pattern for fish: all five endpoints

6.3 Combined Benthic Invertebrate and Fish Responses

Ordination of the output from the meta-analyses made it feasible, for the first time, to quantitatively integrate the results from the benthic invertebrate and fish population surveys. This was made possible by putting the results for these two core EEM components into a common currency, the Hedges' d standardized effect size.

Figure 6-5 shows the multidimensional scaling ordination for the two benthic and three fish endpoints used most often for interpreting effects. The direction of the principal axis correlation vectors was used to divide the ordination space into sectors representing the major categories of effects, in the same way as described in Sections 6.1 and 6.2. In this particular ordination, it was not possible to define sectors for overall nutrient enrichment or limitation (increases in all or decreases in all of the gonad, liver, and condition endpoints) in fish, due to the gonad vector falling in the opposite direction to the liver and condition vectors. Mills falling into these two categories were present in the ordination, but ordinating this particular combination of endpoints did not position these mills into discretely definable sectors.

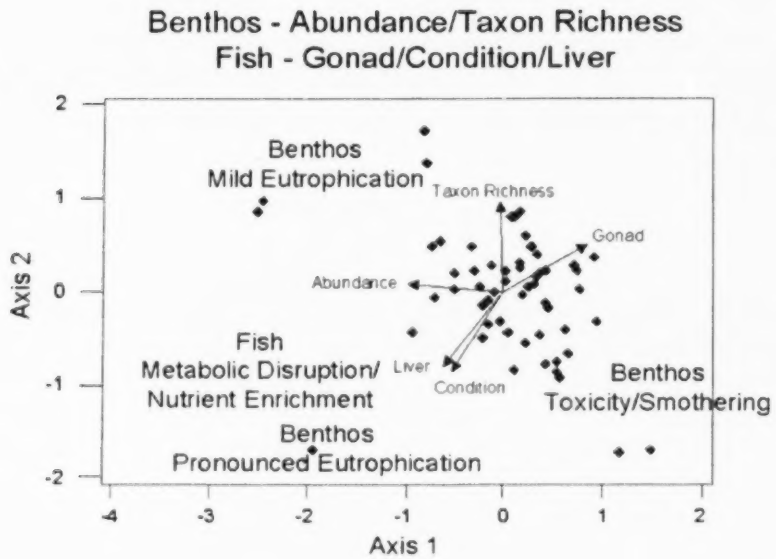


Figure 6-1: Overall national response pattern for benthic invertebrates and fish combined: two benthic and three fish endpoints

Combining the results of the benthic invertebrate and fish population surveys revealed some key findings. Mill exposure areas towards the left side of the ordination space generally showed signs of stimulatory or eutrophication effects, while those towards the right side showed signs of more inhibitory effects. Moving counterclockwise from mills in the upper left sector shows a progression from mild eutrophication to more pronounced eutrophication for mills in the lower left sector, and finally to toxicity/smothering-type effects in the lower right sector. Of note, the fish metabolic disruption/nutrient enrichment sector fell on the benthic eutrophication side of the ordination space, showing good agreement between the results of the benthic invertebrate and fish population surveys.

Inclusion of all four benthic invertebrate and five fish endpoints in one ordination revealed even more key patterns (Fig. 6-6). Collapsing nine dimensions into two resulted in some distortion, but, as discussed above, this was within acceptable limits, and it was possible to approximately locate the sectors for several of the main response patterns by using the positioning of the principal axis correlation vectors. In general, mills in the left and upper portions of the ordination space were associated with more stimulatory/eutrophication-type effects, while those in the right and lower portions were associated with more inhibitory-type effects. Moving counterclockwise

from the 12 o'clock position shows a progression from more pronounced eutrophication through to mild eutrophication in the lower left, to fish nutrient limitation/chemical toxicity in the lower right, to benthic toxicity/smothering in the upper right.

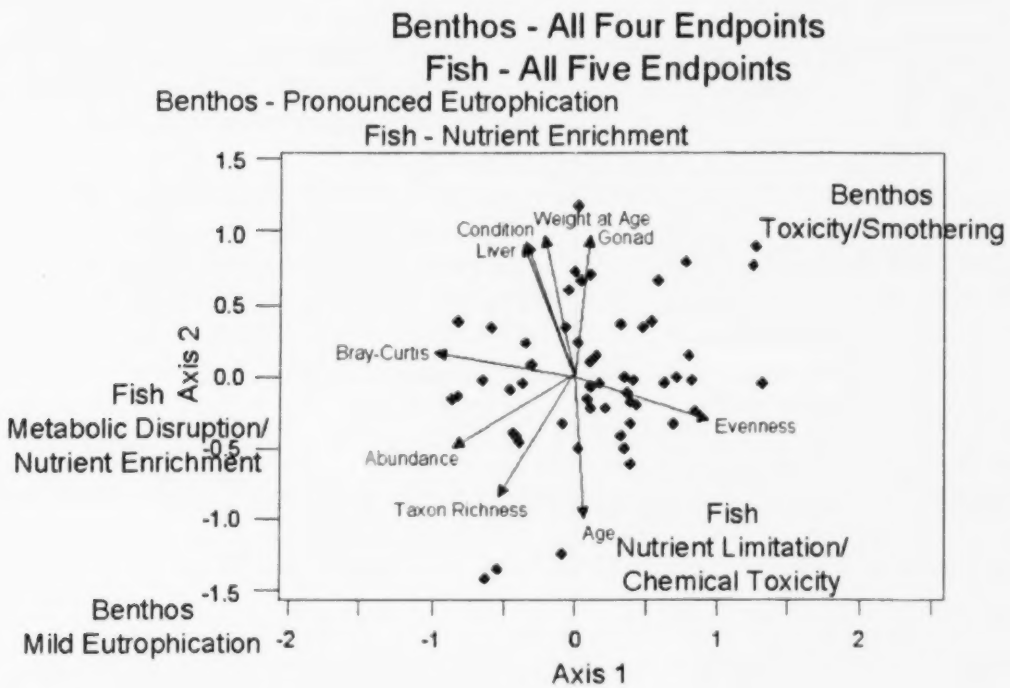


Figure 6-2: Overall national response pattern for benthic invertebrates and fish combined: all benthic and fish endpoints

There was doubtless some overlap among the different sectors, but the overall pattern helps to interpret the benthic invertebrate and fish responses in an integrated fashion. As for Figure 6-5, this figure shows the good agreement between the benthic invertebrate and fish results. As expected, the fish metabolic disruption/nutrient enrichment sectors overlap the benthic invertebrate eutrophication sectors. Interestingly, overall fish nutrient enrichment (without measurable metabolic disruption) may not be manifest until the benthos reach a state of pronounced eutrophication (approximate 12 o'clock position). It is possible that metabolic disruption is masked at high levels of eutrophication. A similar masking effect has also been described for invertebrates (Lowell et al., 2000; Fig. 4-3). On the other hand, under conditions of less pronounced eutrophication, the fish results indicate the presence of metabolic disruption effects, even when the overall measured invertebrate community effect is one of eutrophication. Thus, the two core components of the EEM field survey (benthos and fish) complement each other by revealing different aspects of the effects associated with effluent exposure.

The complementary nature of the nine core benthic invertebrate and fish endpoints is also illustrated by the overall positioning of the vectors, which basically radiate out like the spokes of a wheel (Fig. 6-6). The liver and condition vectors overlap at this scale, but Figure 6-3 shows that these two endpoints do have value for distinguishing different effects at a finer scale. Thus, the preceding figures suggest minimal redundancy in the nine core endpoints. That is, the different endpoints appear to be measuring different (and partially independent) aspects of the complex of effects that pulp mill effluent can have on receiving water biota. As discussed in Sections 6.1 and 6.2, this indicates that each of the endpoints has value, and, at present, it is recommended that all be retained.

7. Interpretation of Sublethal Toxicity Data for Cycle 2 of the Pulp and Paper EEM Program

7.1 Introduction

Under the EEM component of the *Pulp and Paper Effluent Regulations*, each Canadian pulp and paper mill is required to measure the sublethal toxicity of their final effluent discharge. During Cycle 1, a battery of three sublethal toxicity tests was used during four consecutive quarters within the principal year of field work to assess effects on 1) early life stage development of fish, 2) reproduction of an invertebrate, and 3) growth of an aquatic plant. During Cycle 2, three sublethal toxicity tests (fish, invertebrate, and plant) were conducted twice a year (summer and winter) for each of the three years in the cycle. The choice of tests and species depended on whether the mill discharged to a fresh or estuarine/marine water body. Data were reported to government within 90 days and were interpreted at the end of each three-year cycle, along with results of field monitoring (Environment Canada, 1997a).

Sublethal toxicity tests complement the field components of the EEM studies, as the results are interpreted by a "weight-of-evidence" approach using a suite of laboratory and field monitoring tools. Within EEM, the results of sublethal tests on effluent or receiving water can be used to:

- measure changes in toxicity over time, as a result of changes in mill process or effluent treatment;
- estimate the potential for effects in the receiving water; and
- estimate, in multiple-discharge situations, the relative contributions from various sources to any observed effects in the receiving environment.

Environment Canada has prepared written technical guidance to explain how the mills' consultants or government regulators can use sublethal toxicity data to address each of the above purposes (Environment Canada, 1998b, 1999a). This summary gives some examples of these applications, using sublethal toxicity data generated in Cycle 2 of EEM.

7.2 Sublethal Toxicity Test Methods

The specific test methods for Cycle 2 of EEM for the Canadian pulp and paper industry are listed in Table 7-1. For each of fish, invertebrates, and plants, there are freshwater and marine options, and in some categories, there is a choice of species. Prescribed methods for some of the tests have been published by Environment Canada (last column of Table 7-1). Each method document lists the minimum reporting requirements in section 8 or 9; the toxicity data must be accompanied by descriptions of the materials, methods, and calculations. Some of the methods prescribed in Table 7-1 were published by the U.S. Environmental Protection Agency (USEPA, 1994; Chapman et al., 1995). Although section 10 of the EPA documents describes the information to be reported, Environment Canada established minimum requirements for these methods when used in EEM.

Table 7-1: Sublethal toxicity tests for EEM

Description of Test	Receiving Environment	Test Species	Acceptable Method(s)
Early life stage development of fish	Marine/estuarine	Inland silverside or topsmelt	USEPA, 1994; Chapman et al., 1995
	Freshwater	Fathead minnow or salmonid (rainbow trout)	Environment Canada, 1992b, 1998a
Reproduction of an invertebrate	Marine/estuarine	Echinoid (sea urchin or sand dollar)	Environment Canada, 1992d
	Freshwater	*Ceriodaphnia dubia* (water flea)	Environment Canada, 1992a
Toxicity to an aquatic plant	Marine/estuarine	*Champia parvula* (alga)	USEPA, 1994
	Freshwater	*Selenastrum capricornutum* (alga)	Environment Canada, 1992c

To assist in the submission of valid and high-quality data, laboratories that carried out sublethal toxicity testing were required to follow specific conditions for quality assurance/quality control. These conditions were outlined by Environment Canada (1997a, 1998b) and in the specific test method documents. Regional staff of Environment Canada made quality assurance checks upon receipt of test data, before tabulating them.

For most tests, the endpoint that was used to measure effluent quality was the *inhibiting concentration for 25% effect (IC25)*, i.e., that concentration causing performance that was 25% inferior to that of the control organisms. The endpoint for the rainbow trout embryo test was the *effective concentration for 25% of the embryos (EC25)*. In a freshwater test, if full-strength effluent did not cause 25% inhibition/effect, then the endpoint was reported as >100% concentration. For estuarine/marine tests, lack of effect was reported as >67% to >71%, which was the highest concentration that could be tested due to the standard procedure of adding hypersaline brine to the effluent for salinity adjustment.

7.3 Monitoring Changes in Effluent Quality Between Cycles

Sublethal toxicity endpoints can be used to compare the quality of effluent at different times. Comparisons between results for Cycles 1 and 2 were made for the complete distributions of endpoints at all Canadian mills. Such a comparison was made for each freshwater and marine test species, as illustrated in Figures 7-1 and 7-2. Comparisons were based on the number of IC25s falling into each category range (i.e., ≥100%, ≥60% to <100%, ≥36% to <60%, etc.). The highest concentration range for effluents that were salinity-adjusted has been summarized in a category designated =60%.

A definite trend towards improved effluent quality was shown by comparisons between Cycles 1 and 2, for the complete data sets of each freshwater and marine test species (Figs. 7-1 and 7-2). The toxic concentrations reported by all mills are shown in the figures. The national trend for improvement was particularly clear in toxicity levels for *Ceriodaphnia*, fathead minnow, *Selenastrum*, and inland silverside (Figs. 7-1A, 7-1B, 7-1D, and 7-2B). For those species, toxic levels for Cycle 1 tended to be distributed over a wide range of concentrations, including some low concentrations of 1 or 2% for some mills. There was only a slight tendency for higher clusters of "no-effect" results at the left side of each histogram. In Cycle 2, the histograms for these same four species show a distribution at higher concentrations (less toxic), with marked increases in no-effect results.

Results for tests with echinoderms and *Champia parvula* show less striking tendency for lowered toxicity in Cycle 2. No comparisons between Cycles 1 and 2 could be made for the topsmelt growth inhibition test (Figs. 7-2C and 7-3B), since the test was not used in Cycle 1.

The opposite trend was seen in the rainbow trout embryo test, with an apparent slight worsening of effluent quality (Fig. 7-1C). This is likely because the test method for rainbow trout was substantially improved after the round of testing in Cycle 1, resulting in a more responsive test.

A clear national improvement in effluent quality between Cycles 1 and 2 was also shown by comparing the percentage of tests with IC25s showing no sublethal response in the highest test concentration. This comparison was also done individually for each test species and is based on all the endpoints submitted by all mills in a given cycle for a given species (i.e., all tests with a given species, from all mills, that showed IC25 >100% for freshwater tests or >60% for marine tests).

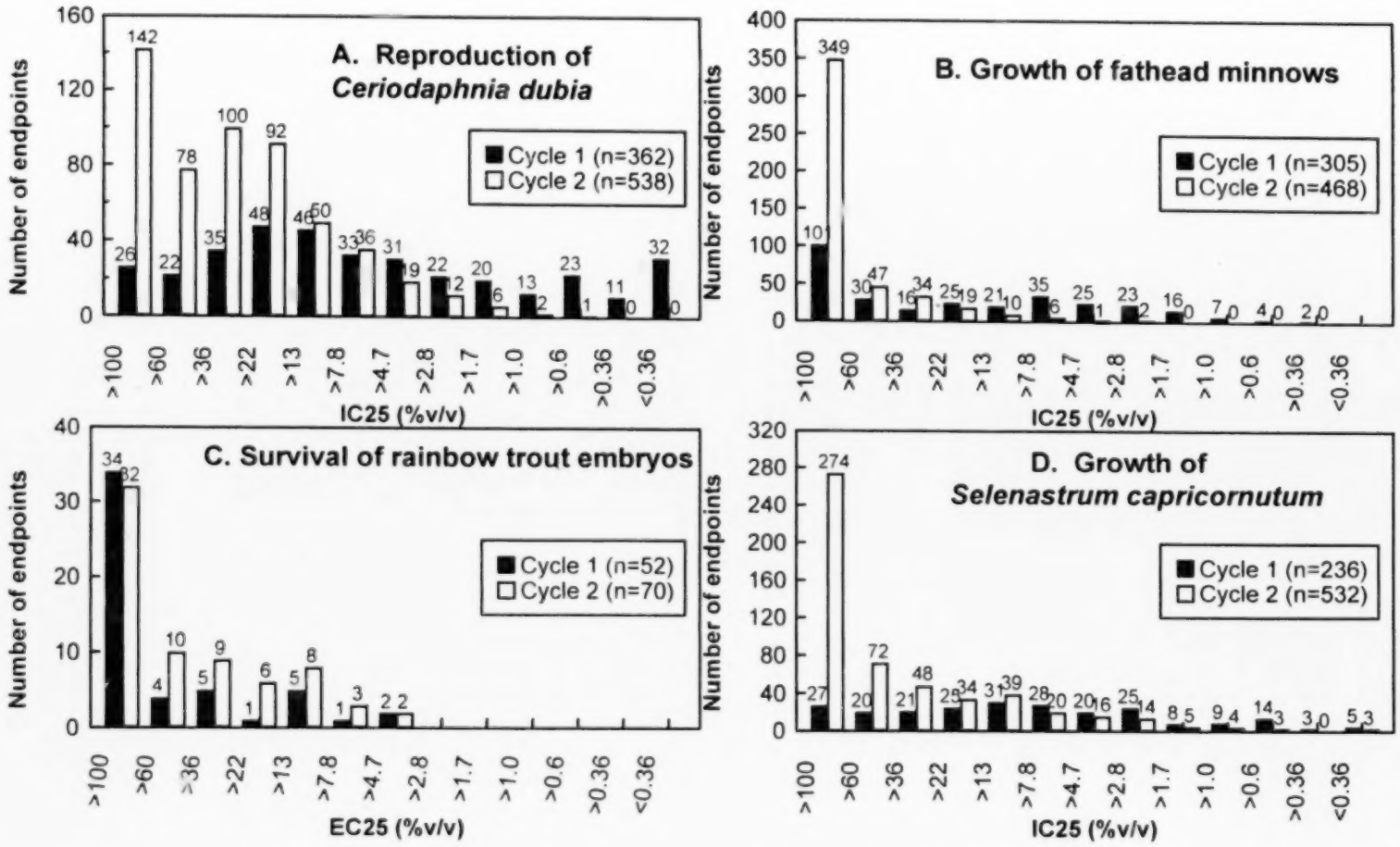


Figure 7-1: Comparison of sublethal toxicity to freshwater species in Cycles 1 and 2 of EEM

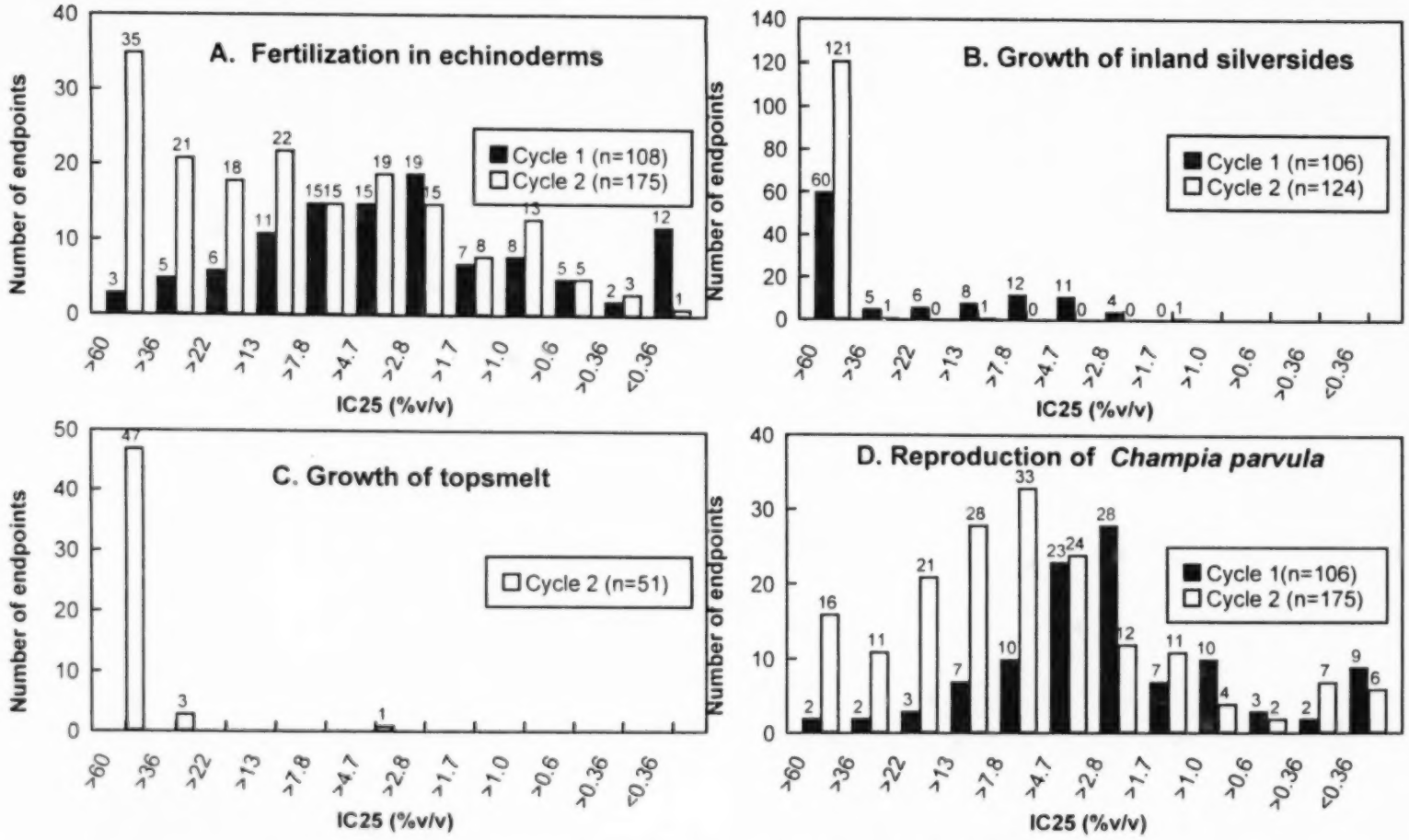


Figure 7-2: Comparison of sublethal toxicity to marine species in Cycles 1 and 2 of EEM

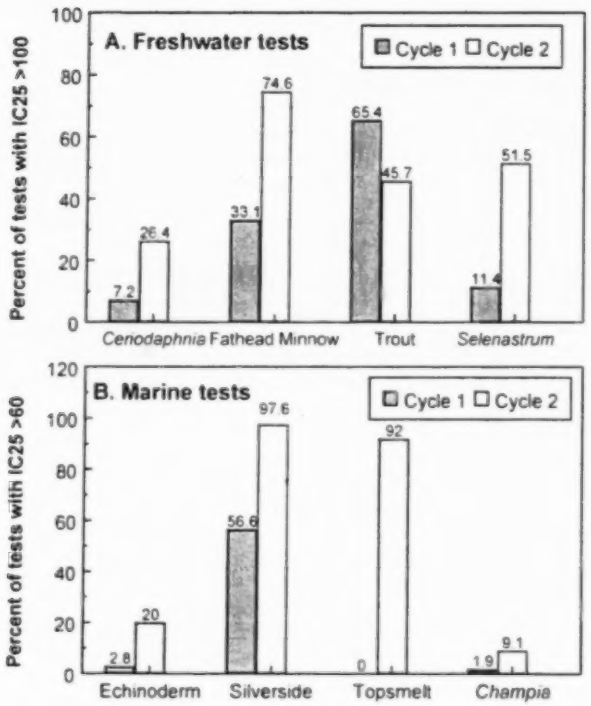


Figure 7-3: Percentages of sublethal tests showing no effect at highest concentration

This trend was again most pronounced in tests with fathead minnow, *Selenastrum*, and inland silverside (Figs. 7-3A and 7-3B).

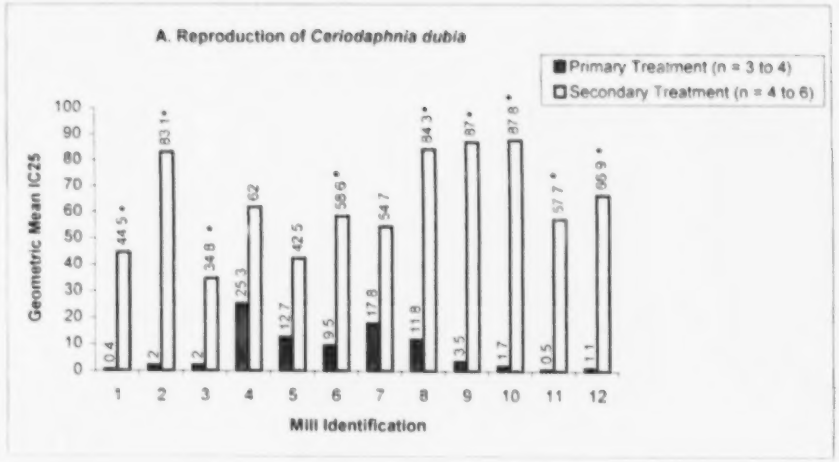
The trend for improvement was also evident in the percentages of non-responsive tests with invertebrates (*Ceriodaphnia* and echinoderms) and the alga *Champia*, but was less striking because percentages were relatively low in both cycles. The perception of smaller improvements shown by the echinoderm and *Champia* tests (Fig. 7-3B) might also be partly due to changes in the method of adjusting salinity of effluents. In Cycle 1, dry salts were used to adjust salinity. In Cycle 2, hypersaline brine was used, which diluted effluents by 29–33% (Environment Canada, 1997b). Therefore, toxicity could not be determined at effluent concentrations above 67–71%.

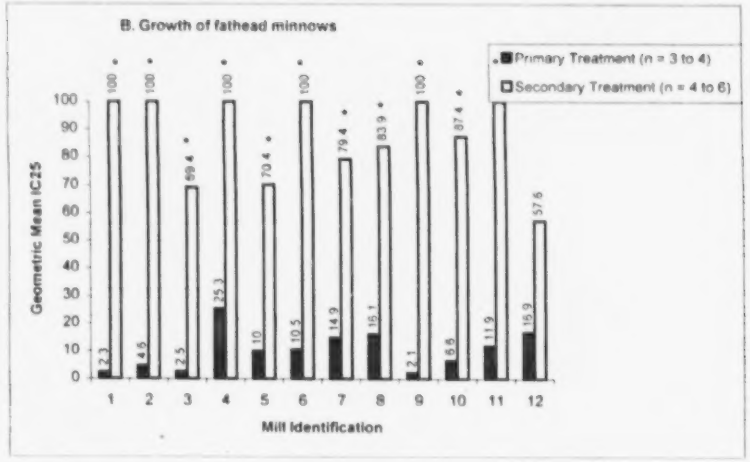
The test of embryo viability for rainbow trout was the only one showing the opposite trend (Fig. 7-3A). Again, that was most likely the result of the improved method for performing this test in Cycle 2. The topsmelt had no data for Cycle 1, and hence no comparison. Every other species showed improvement of effluent over the time period.

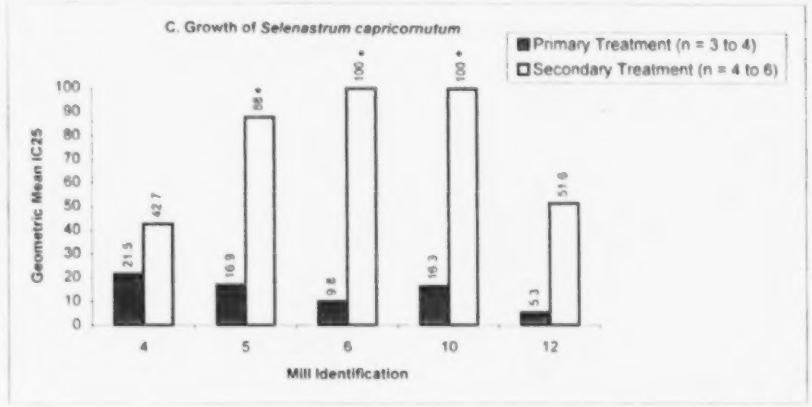
7.4 Monitoring Changes with Treatment of Effluent

Changes in Cycle 1 and Cycle 2 effluent quality resulting from the installation of new secondary treatment systems were determined by comparing IC25s from Ontario Region mills that did not have secondary treatment during Cycle 1, but did have it during Cycle 2. The geometric mean IC25 for primary-treated effluent was compared with the geometric mean IC25 after treatment was installed. This was done for each species that had sufficient data, as illustrated in Figure 7-4. A similar comparison of geometric means evaluated temporal changes in effluent quality at freshwater mills in British Columbia that had secondary treatment before the beginning of Cycle 1. This is illustrated in Figure 7-5.

In calculating the geometric mean IC25s, a test that did not cause 25% sublethal effect at any concentration including 100% or resulted in an IC25 of 100% was assigned, for purposes of calculation, an endpoint of 100. Although this action might seem to be biased against the industry being evaluated, that is really not the case. Assigning a value of 100% almost always means that this hypothetical IC25 is closer to the other measured IC25s, reducing the standard deviation among IC25s. The benefit to the company would be the calculation of a less toxic value for the *predicted minimum toxic concentration* (see below; an estimate of the most toxic level expected during extensive testing of a given effluent). Standard deviations were calculated using the logarithms of the IC25s (EC25s for rainbow trout).







* An asterisk at the top of a column indicates that results for secondary treatment were significantly different from those for primary treatment at the same mill (Mann-Whitney U test, $P<0.05$)

Figure 7-1: Sublethal toxicity of Ontario mill effluents before secondary treatment (Cycle 1 of EEM) and after (Cycle 2)

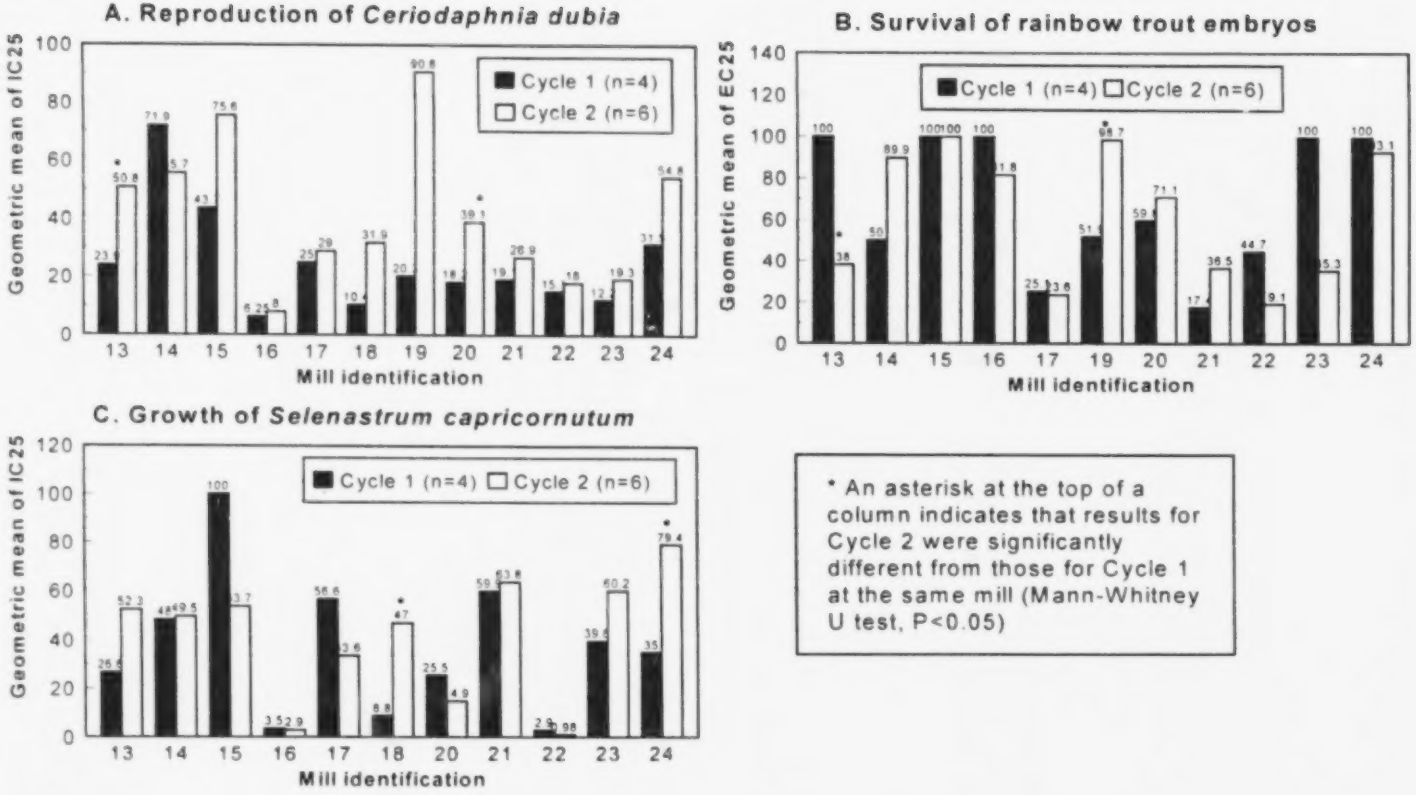

A. Reproduction of Ceriodaphnia dubia
Geometric mean of IC25
Cycle 1 (n=4)
Cycle 2 (n=6)
Mill identification
B. Survival of rainbow trout embryos
Geometric mean of EC25
Cycle 1 (n=4) Cycle 2 (n=6)
Mill identification
C. Growth of Selenastrum capricornutum
Geometric mean of IC25
Cycle 1 (n=4) Cycle 2 (n=6)
Mill identification
* An asterisk at the top of a column indicates that results for Cycle 2 were significantly different from those for Cycle 1 at the same mill (Mann-Whitney U test, P<0.05)

Figure 7-2: Comparison of sublethal toxicity of secondary-treated effluent from freshwater mills in British Columbia in Cycles 1 and 2 of EEM

EEM data from the Ontario Region showed general improvements in effluent quality based on the geometric mean IC25s. It was very evident that toxicity decreased as mills changed from primary waste treatment to secondary treatment. The convincing improvement in effluent quality is shown in Figure 7-4; in most cases, the corresponding IC25 was higher for secondary treatment.

Data from the Pacific and Yukon Region indicate that secondary treatment, not other temporal changes between Cycles 1 and 2, was the important factor for improved effluents. Temporal changes in effluent quality were much less pronounced in effluents from British Columbia mills that had used secondary treatment during both Cycles 1 and 2 (Fig. 7-5).

7.5 Conclusions

Actual situations of discharge from pulp and paper mills, evaluated in the EEM program, clearly illustrate the application of sublethal toxicity tests to measure temporal changes in mill process or effluent treatment, to estimate potential effects in the receiving water, and to estimate relative contributions from individual sources of a multiple-discharge situation.

The data for sublethal toxicity tests summarized here show a national improvement in effluent quality at Canadian pulp and paper mills between Cycle 1 and Cycle 2. Moreover, the effectiveness of sublethal toxicity tests in identifying a trend in effluent quality is evident. The same degree of effectiveness was seen by others, in data from Cycle 1 of EEM for the Canadian pulp and paper industry (Environment Canada, 1997a). Between 1992 and 1996, many Canadian mills upgraded effluent treatment systems or built new components to meet requirements for TSS, BOD, and the absence of acute lethality to rainbow trout. In many cases, as a new or upgraded treatment system became operational, a decrease in sublethal toxicity was evident in the next period of effluent testing. Combining results for all effluents receiving secondary treatment showed clearly lowered toxicity compared with effluents having only primary treatment or recently implemented secondary treatment (Environment Canada, 1997a).

8. Summary and Conclusions

The purpose of this report was to assess the current state of the aquatic environment around pulp and paper mills in Canada based on the detailed analysis of data collected in the National EEM Program since 1992 and to present any discernible trends in the receiving environment over the past 10 years. The assessment has provided a broad picture of what types of effects are occurring in the aquatic receiving environment, the magnitude of these effects, and whether there are any observable patterns linking the effects to various mill processes and effluent treatment types.

8.1 Historical Context

The pulp and paper industry has benefited the Canadian economy in many ways while at the same time pulp and paper mill effluents have long been the source of environmental concern. The effluents became the focus of attention following a series of studies in the late 1980's demonstrating a number of impacts on fish located near pulp and paper mills. These changes included abnormal size distribution and age structure, potential impacts on the reproductive performance of fish (e.g., gonad size, age to maturity), changes in recruitment and mortality, and imbalances in fish community structures. Effluents were also showing impacts on the downstream structure and functioning of benthic invertebrate communities, which often exhibited fewer species, lower densities, and altered taxonomic composition compared with benthic invertebrate communities at a reference site.

Although it was recognized that each pulp and paper mill effluent exhibits varying characteristics, depending on a number of factors, three conventional pollutants were implicated in these adverse environmental impacts — fibre and suspended solids, colour and turbidity, and organic and nutrient enrichment loads. Pulp and paper mills were also identified as the main industrial source of dioxins and furans, which, following their release in the effluent, were highly persistent in the environment and acutely toxic to aquatic organisms.

As a result of environmental concern about the adverse impacts of pulp mill effluents on aquatic ecosystems, regulatory actions were taken in order to reduce effluent toxicity. In 1992, the federal government passed the Pulp and Paper Regulatory Framework consisting of the *Pulp and Paper Effluent Regulations* (PPER) under the *Fisheries Act* and two regulations under the *Canadian Environmental Protection Act*. The PPER revoked and replaced an earlier set of regulations passed in 1971. The PPER established discharge limits for BOD, TSS and prohibit the discharge of effluent that are acute lethal to rainbow trout. The CEPA regulations were designed to ensure that chlorinated dioxins and furans were not formed during pulp bleaching.

The implementation of these new regulations led to major capital investment in the effluent treatment processes at pulp and paper mills, which significantly improved the quality of the mill effluents. However, it was recognized that these measures alone might not ensure adequate protection of the aquatic ecosystem at every site. It was also recognized that the assessment of the environmental impact of pulp and paper mill effluents was a difficult task, as a result of the drastic variability in the chemical composition of pulp and paper mill effluents as well as the complexity and natural variability of biological systems. Consequently, the 1992 PPER included the addition of a requirement for an EEM program in order to provide an overview of the state of

receiving environments around pulp and paper mills across Canada, to attempt to demonstrate whether or not the major improvements in effluent quality resulting from improvements in treatment processes were associated with similar improvements in the receiving environment, and to evaluate the long-term effects of pulp and paper mill effluents on Canadian aquatic ecosystems.

8.2 Components of the National EEM Program

The National EEM Program requires Canada's pulp and paper mills to conduct studies on their receiving environments in order to assess and monitor the potential effects of their effluents on fish, fish habitat, and the use of fisheries resources by humans. The program is structured in a three- or four-year sequence of monitoring and interpretation phases known as "cycles."

There are 134 pulp and paper mills currently conducting EEM in Canada. Each mill is expected to develop a site-specific monitoring program in collaboration with regional Environment Canada officials at the beginning of each cycle. At the end of each cycle, mills are required to submit an interpretative report summarizing their field work and interpreting their results.

An EEM study includes the following components:

- a fish population survey to assess the health of fish;
- a benthic invertebrate community survey to assess the fish habitat;
- a study of dioxins and furans in edible fish tissues and a tainting study as assessments of the usability of fisheries resources;
- sublethal toxicity testing to assess effluent quality; and
- supporting water and sediment quality variables to aid in the interpretation of biological data.

Information from EEM, taken together with social, economic, and technical considerations, can eventually be used to assess the adequacy of the *Pulp and Paper Effluent Regulations* in protecting fish populations, fish habitat, and the use of fisheries resources and to indicate where there is a local, regional, or national need for further protection.

8.2.1 Field Study Components

The benthic invertebrate community survey and the fish population survey are highly complementary and together form the basis of a solid EEM study.

An effect in the EEM is defined as a statistically significant difference in measurements between an exposure area and a reference area. As such, an increase or a decrease in the selected effect endpoints in either of these two surveys constitutes an effect.

8.2.1.1 Benthic Invertebrate Community Survey

Benthic invertebrate communities are excellent indicators of the health of fish habitats, and variations in their structure and function are signs of habitat degradation due to contamination. The release of large volumes of organic matter in pulp and paper mill effluents can significantly

alter the receiving aquatic environment and lead to increases in nutrient enrichment, primary production, and toxic inhibitions, all of which can have detrimental impacts on benthic invertebrate communities. The degradation of this organic matter in the aquatic environment can also create anoxic conditions and smother the substrate, thereby also impacting the invertebrate community.

Anoxic conditions have been associated with lower invertebrate feeding rates, potentially lowering their growth and biomass. Conversely, studies have also shown increases in abundance, growth, survival, and reproductive output for invertebrates downstream of pulp and paper mills discharging substantial quantities of organic matter and nutrients.

The EEM benthic invertebrate community survey endpoints are total abundance (of individual benthic invertebrates), taxon richness (number of taxa), Bray-Curtis index of dissimilarity (change in overall community structure, particularly community composition), and Simpson's evenness (how evenly individuals are distributed among the taxa). Changes in exposure areas relative to reference areas in these four core endpoints were measured to evaluate benthic community structure and its functioning. Specifically, these endpoints were used to determine the magnitude and spatial extent of habitat degradation (due to organic enrichment or other forms of physical and chemical contamination) and to provide information on energy available to fish selected as sentinel species in the fish population survey and on changes in the types of organisms present in exposure and reference areas.

8.2.1.2 Fish Population Survey

The approach of the fish population survey is based on measuring the growth, reproduction, and survival of fish species exposed to effluent and comparing the results with those of the same fish species collected in a reference area. The fish population survey is based on the assumption that looking at the health of one or two fish species will provide information concerning the fish community itself and possibly provide some insight with regard to other aspects of the aquatic ecosystem.

The EEM fish population survey endpoints used to assess the potential effects of effluent on fish are gonad weight, condition factor, liver weight, weight at age, and age. These can be grouped according to summary categories of age structure, energy expenditure (gonad weight), and energy storage (condition factor and liver weight). Weight at age is a measure of the growth rate described by the relationship of size (weight) to age; generally, the rate of increase in weight declines as fish approach the limit of their life span.

Response characteristics of fish in the exposure area are identified by an increase (+), a decrease (-), or no difference (0) compared with fish in the reference area. Generalized response patterns can then be identified that are a useful tool in determining what form subsequent monitoring should take, identifying the cause of the effects, and determining follow-up action.

In Cycle 2, alternative monitoring methods included on-site flow-through fathead minnow bioassays, mesocosm studies, and the use of caged bivalves. Different endpoints were used for

those marine mills that chose bivalves for the EEM program, including size, shell weight or density, age, and external abnormalities.

8.2.2 Sublethal Toxicity Testing Component

Under EEM, each Canadian pulp and paper mill is required to measure the sublethal toxicity of their final effluent discharge. In Cycle 2, three sublethal toxicity tests (early life stage development of fish, reproduction of an invertebrate, and toxicity to an aquatic plant) were conducted twice a year (summer and winter) for each of the three years in the cycle. The choice of test and species depended on whether the mill discharged to a fresh or estuarine/marine water body.

Sublethal toxicity tests complement the field components of the EEM studies. Within EEM, the results of sublethal toxicity tests on effluent or receiving water can be used to measure changes in toxicity over time as a result of changes in mill process or effluent treatment, to estimate the potential for effects in the receiving water, and to estimate, in multiple-discharge situations, the relative contributions from various sources to any observed effects in the receiving environment.

8.3 Cycle 2 Results

Cycle 1 data were used as a baseline to gain a better understanding of the variability of the field measurements. This led to an improvement of Cycle 2 study designs and gave some insight into the typical response patterns of fish and benthos to pulp mill effluents. The focus of this national assessment was placed on Cycle 2 data, with some comparisons with Cycle 1 data.

8.3.1 Benthic Invertebrate Community Survey

About two-thirds of the interpretative reports submitted by each mill reported notable changes in benthic conditions between cycles. Of the reports that noted changes, just under three-quarters of the studies reported an improvement in benthic conditions between cycles. This overall improvement was reported across the country, although there were some differences among regions. A smaller number of reports indicated a worsening of benthic conditions, although some of these cases may have been due to changes in other non-mill-related factors. In many cases, the improvements in benthic conditions were attributed to improvements in mill effluent treatment practices, which led to significant decreases in effluent sublethal toxicity and lessened eutrophication.

Inspection of the grand means from the meta-analyses for mills using a Control/Impact study design (magnitude of effects measured as proportional differences between reference and exposure areas) revealed that a particularly common effect associated with exposure to effluent was an increase in invertebrate numbers (abundance). This is a well-known eutrophication-type response to exposure to pulp mill effluent, due to the nutrient enrichment effects of the phosphorus, nitrogen, and organic content of the effluent. In contrast, there was a more even distribution of mills with respect to increases, decreases, and no significant changes in number of taxa (taxon richness). These two findings, taken together, are indicative of mild eutrophication.

The small number of mills with increases in abundance and decreases in taxon richness were showing typical signs of more pronounced eutrophication.

There was also a fairly even distribution of mills with respect to increases, decreases, or no significant changes in evenness. On the other hand, the Bray-Curtis grand mean was highly significantly positive, showing that there were clear changes in benthic invertebrate community composition in exposure areas across the country.

When the results from the Gradient design mills (magnitude of effects determined by regressing endpoints against distance from mill outfall and calculating the correlation coefficient) were combined with those from the Control/Impact design mills, the overall results were similar to those for Control/Impact mills alone, again suggesting an overall predominance of mild to moderate eutrophication effects.

A general appreciation of the overall magnitudes of measured effects was also conveyed by plotting histograms of non-standardized effect sizes for each endpoint (effect size calculated as the percent change in effect size relative to the reference mean). The abundance endpoint showed the most extreme range in exposure-associated effects (decreases of almost 100% to increases of over 2000%). Measured effects for taxon richness ranged from decreases of 70% to increases of 150%. Evenness effect sizes varied from decreases of just over 60% to increases of 190%. Bray-Curtis effect sizes ranged from approximately zero to increases of up to 500%. Thus, the magnitudes of effects associated with effluent exposure were often quite large, although these magnitudes should be considered within the context of the range of variability typically seen in reference areas (which was accounted for in the standardized effect sizes used for the meta-analyses).

8.3.1.1 Habitat Type

The magnitude and direction of effluent-associated effects on benthic invertebrate communities varied significantly among different habitat types for both Control/Impact and Gradient mills. Abundance was significantly increased in both erosional and depositional riverine habitats, whereas the distribution of mill responses for taxon richness mostly ranged from increases to no significant change. This combination of responses is indicative of an average effect of mild eutrophication in river exposure areas. In contrast, both abundance and taxon richness decreased in marine/estuarine habitats, a response that is indicative of an average toxicity or smothering effect. The distribution for lacustrine habitats was intermediate between those for riverine and marine/estuarine habitats.

Abundance effects in estuarine habitats were intermediate between those in freshwater and marine habitats, in keeping with the intermediate physical characteristics of estuarine habitats. The inhibitory response seen in estuarine and marine habitats is likely due to a combination of factors, including the location of sampling stations close to the outfall and habitat deterioration due to historical deposition of thick fibre mats.

In erosional riverine habitats, evenness was nearly significantly reduced overall in exposure areas, which may reflect the sensitivity of organisms in erosional habitats to even mild

eutrophication. In contrast, evenness was significantly increased in estuarine exposure areas (and nearly so for marine exposure areas), suggesting that toxicity/smothering effects may have led to the loss of more rare taxa, leaving behind a smaller number of more evenly distributed, pollution-tolerant taxa.

The largest exposure area effects for the habitat subgroups were observed for the sensitive Bray-Curtis endpoint. By far the largest exposure-associated changes in community composition were measured in lacustrine habitats, in contrast with other endpoints, where the measured lake effects were more intermediate. This illustrates that, under different environmental conditions, effluent-associated effects may be manifested in different ways, and it underscores the value of measuring changes in community structure using multiple endpoints to reduce the probability of missing important effects.

8.3.1.2 Effluent Treatment Type

Exposure area abundance was significantly increased for the two most common effluent treatment types — activated sludge treatment and aerated stabilization basin. The magnitude of this effect was quite similar between the two types. There was also a high degree of overlap for taxon richness. Thus, the response pattern for these two prevalent treatment types was similar and reflected the mild eutrophication effect that was most commonly observed for benthic invertebrate communities in Cycle 2.

The two major effluent treatment types in Control/Impact mills were also quite similar in terms of effects on evenness (not significantly different from zero). Adding in the Gradient mills resulted in increased eveness, probably reflecting the effect seen in marine and estuarine mills. The magnitudes of effects were larger for the Bray-Curtis endpoint, and these measured changes in community composition were fairly similar between the two treatment types.

8.3.1.3 Mill Process Type

As was found for effluent treatment, exposure area abundance was significantly increased to a similar degree for the two most common mill process types — kraft and thermomechanical. The two types differed for the taxon richness endpoint, though, with the results suggesting a milder eutrophication response for kraft mills (more typical of the national norm) and a more pronounced eutrophication response for thermomechanical mills.

The two mill process-type responses were fairly similar for the evenness endpoint responses. The results indicated that individual mills that had either significant increases or decreases in evenness could not be distinguished on the basis of mill process type. The magnitude of effect was greater for the Bray-Curtis endpoint, but the two mill process types did not significantly differ.

8.3.2 Fish Population Survey

In Cycle 2, 114 interpretative reports contained fish population survey information. The results presented in this report focused on the analysis of the electronic data, as the mill-reported results

have been published elsewhere. The electronic data submitted to Environment Canada by mills were used to conduct ANCOVA/ANOVA analyses in conjunction with meta-analyses, and, where applicable, results were compared with those obtained from the mill interpretative reports. Generally, there was good agreement between the results submitted by mills and the analyses carried out on the electronic data.

The incidence of effects was consistent between male and female fish, and results were similar between fish species within sites. Based on ANCOVA analyses, nationally, the overall median response was an increase in condition factor and liver weight and a decrease in gonad weight in fish exposed to effluent. An increase in condition factor was found more than twice as often as a decrease. For liver weight, an increase was found rather than a decrease in 40% of the cases. A decrease in gonad weight was found twice as often as an increase. In approximately 10% of the ANCOVA analyses for condition factor, liver weight, and gonad weight, there were statistical interactions. Interactions in the data can be interpreted as actual responses and were not discarded from the analysis.

To provide a general picture of the overall magnitude of effects, non-standardized effect sizes within the exposure areas were illustrated by plotting percent changes in condition factor, gonad weight, and liver weight (excluding data with a significant interaction) (see Section 8.3.1 and Chapters 3–5 for further discussion of standardized versus non-standardized effect sizes). For condition factor, the percent change varied from a decrease of 27% to an increase of 24%. For gonad weight, the percent change ranged from a decrease of 90% to an increase of 286%. The percent change in liver weight ranged from a decrease of 88% to an increase of 73%. On average (calculated with sign, or direction of effect, included for each mill), a 2% increase in condition factor, a 5% decrease in gonad weight, a 3% increase in liver weight, and a 5% increase in weight at age were found.

On a national level, it appears that the fish downstream of pulp mill effluents are experiencing a combination of metabolic disruption (a change in the ability to process energy, indicated by an increase in condition factor and liver weight and a decrease in gonad weight) and nutrient enrichment (an increase in food availability, indicated by an increase in liver weight, gonad weight, and condition factor). The general pattern that was found in Cycle 2 was consistent with that of Cycle 1.

Using the electronic data, the percentages of fish comparisons where the measured effect size exceeded two standard deviations, 25%, and 10% were examined. The Cycle 2 data showed that an effect size of 20–30% for gonad weight, as well as liver weight and weight at age, was appropriate for determining sample sizes *a priori* in EEM. An effect size of 10% proved to be more appropriate for condition factor.

Cycle 2 data were analyzed using meta-analysis to determine the standardized magnitude of effects on the key fish endpoints as influenced by gender, fish species, mill process type, effluent treatment type, and habitat of the receiving area.

8.3.2.1 Gender

Male and female fish had the same responses in all response variables except age, which was increased in both males and females, but not significantly so in males. There was a significant decrease in gonad weight in fish from exposure areas, with the decrease being significantly greater for females than for males.

8.3.2.2 Fish Species

The analysis using species as a grouping variable was conducted on 19 fish species. The effects found in the two most common species (white sucker and longnose sucker) were consistent with the national grand means for gonad weight, liver weight, condition factor, and weight at age. That is, they revealed a significant increase in liver weight, condition factor, and weight at age and a significant decrease in gonad weight. The longnose sucker also exhibited results consistent with the national grand mean for age (significant increase in the exposure area), but the average age structure for white suckers was not significantly different in the exposure and reference areas.

Trout-perch and yellow perch, which were also commonly monitored in Cycle 2, had unique response patterns. For instance, liver weight was the only endpoint where there was a significant change (i.e., decrease) between exposure and reference areas for trout-perch. The national results found for yellow perch (significantly decreased condition factor, gonad weight, and weight at age and non-significantly decreased liver weight) were consistent with what is typically found in situations of nutrient limitation.

8.3.2.3 Mill Process Type

The two most common mill process types were thermomechanical pulp and kraft. Mills using a thermomechanical process type had significant differences in all five key endpoints that were consistent with the national grand means. Kraft mills were almost identical, with the exception of liver weight, which showed an increase that was not significant compared with the reference area.

8.3.2.4 Effluent Treatment Type

Both activated sludge treatment and aerated stabilization basins followed the national trend, with a statistically significant increase in weight at age, liver weight, and condition and a statistically significant decrease in gonad weight in fish sampled in the exposure area. Age was increased at sites with both treatment types, although this difference was not significant for aerated stabilization basins.

8.3.2.5 Habitat

Rivers were by far the most commonly sampled habitat type for fish in Cycle 2. All three habitat types — rivers, lakes, and marine/estuarine habitats — demonstrated a significant decrease in gonad weight, with fish collected in lakes and marine/estuarine sites showing a greater impact on

gonad weight. Condition was significantly increased for all three habitat types. Liver weight was significantly increased for rivers and marine/estuarine sites, with no significant difference in lakes. Weight at age was significantly increased for rivers and lakes, with no significant difference in marine/estuarine sites. Age was significantly increased for rivers and significantly decreased in the other two habitat types.

Rivers had significant changes in all five response variables, while lakes and marine/estuarine habitats had significant changes in four of five variables.

8.3.3 National Response Patterns for Combined Benthic Invertebrate Community and Fish Population Surveys

The results from the benthic invertebrate and fish population surveys were quantitatively integrated using multidimensional scaling ordination. The overall pattern showed good agreement between the benthic invertebrate and fish results. As expected, the fish metabolic disruption/nutrient enrichment sectors overlapped the benthic invertebrate eutrophication sectors. Interestingly, the results revealed that overall fish nutrient enrichment (without measurable metabolic disruption) may not be manifest until the benthos reach a state of pronounced eutrophication. It is possible that metabolic disruption is masked at high levels of eutrophication. On the other hand, under conditions of less pronounced eutrophication, the fish results indicated the presence of metabolic disruption effects, even when the overall measured invertebrate community effect was one of eutrophication. Thus, the two core components of the EEM field survey complemented each other by revealing different aspects of the effects associated with effluent exposure.

The overall results also suggested minimal redundancy in the nine core endpoints. In other words, the different endpoints appeared to be measuring different (and partially independent) aspects of the complex of effects that pulp mill effluents may have. As a result, each of the endpoints provided a different kind of information that was useful for interpreting the patterns of effects.

8.3.4 Sublethal Toxicity Testing

A definite trend towards improved effluent quality was shown in comparisons between Cycles 1 and 2 for the complete sublethal toxicity data sets of both freshwater and marine test species. The national trend for improvement was particularly clear in sublethal toxicity levels for *Ceriodaphnia*, fathead minnow, *Selenastrum*, and inland silverside. Results for tests with echinoderms and *Champia parvula* showed a less striking tendency for lowered toxicity in Cycle 2. The opposite trend was seen in the rainbow trout embryo tests, with an apparent slight worsening of effluent quality. This was likely because the test method for rainbow trout was substantially improved after Cycle 1, resulting in a more responsive test.

A clear national improvement in effluent quality between Cycles 1 and 2 was also shown by comparing the percentage of tests with IC25s (i.e., the concentrations causing performance that was 25% inferior to that of control organisms) showing no sublethal response in the highest test concentration. The trend was again more pronounced in tests with fathead minnow, *Selenastrum*,

and inland silverside. The trend for improvement was also evident in the percentages of non-responsive tests with invertebrates (*Ceriodaphnia* and echinoderms) and the alga *Champia*, but was less striking, because percentages were relatively low in both cycles. The perception of smaller improvements shown by the echinoderm and *Champia* tests might also be due to changes in the method of adjusting the salinity of effluents. The test of embryo viability for rainbow trout was again the only one showing the opposite trend, probably the result of the improved methods for performing this test in Cycle 2.

The effectiveness of sublethal toxicity tests in identifying a trend in effluent quality is evident from the Cycle 2 analysis.

8.4 Conclusions and Recommendations

8.4.1 Endpoints

8.4.1.1 Benthic Invertebrate Community Survey

Each of the core endpoints used in the benthic invertebrate community survey was chosen to help evaluate a different key aspect of community structure that might be altered by exposure to effluent. The relative sensitivity of the four endpoints was investigated to determine which aspects of community structure show the greatest changes following exposure. By tabulating individual mill results and using meta-analyses, it was shown that the Bray-Curtis endpoint (which is particularly sensitive to changes in community composition) was the most sensitive endpoint. Over half the mills measured a significant change in this endpoint in exposure areas, indicating that a shift in invertebrate community composition is one of the more likely changes to be expected in effluent exposure areas.

Another endpoint sensitivity issue that was evaluated was the relative advantage of using diversity versus evenness as an evaluation endpoint. An analysis of the data for Control/Impact marine/estuarine mills illustrated a pattern that was observed for all the mill groupings — that the diversity endpoint always fell at some point in between taxon richness and evenness (which was expected, given that diversity is a combination of these two endpoints). This showed that reporting diversity (rather than evenness) may yield little more information than what is already captured by taxon richness. In addition, reporting diversity may miss the potential effect of effluent on evenness, which measures an important aspect of community structure that is independent of the information gained from measuring taxon richness. Therefore, reporting evenness probably has greater value than reporting diversity, as long as taxon richness is also reported.

8.4.1.2 Fish Population Survey

The key endpoints used for the evaluation of potential effects of effluent on fish include gonad weight, condition factor, liver weight, weight at age, and age. The measurements of these endpoints help evaluate different key aspects of fish that could be altered by exposure to pulp and paper mill effluent. These results can then be extrapolated to provide insight on the state of the fish population as a whole. The initial analysis of the mill interpretative reports showed that

mills consistently reported on the following three key endpoints: condition factor, gonad weight, and liver weight. As such, the focus of the ANCOVA analysis was on these endpoints, using the following combinations:

- gonad weight as an indicator of energy use, and liver weight and condition factor as indicators of energy storage; or
- gonad weight as an indicator of reproduction, and liver weight and condition factor as indicators of the condition of fish.

Based on the electronic data, less than 5% of mills had no significant differences in each of the three key endpoints examined. More than 40% of mills had at least one statistical difference in all three endpoints. More than half the studies found significant differences in comparisons for condition factor, gonad weight, and liver weight for up to two species of fish and two sexes.

8.4.2 Study Design

8.4.2.1 Benthic Invertebrate Community Survey

The relative advantages of the two basic study designs used in the benthic invertebrate community survey were investigated. It was evident that the different mill groupings were more clearly separated for the Control/Impact designs. It was also shown that fewer significant differences were detected using the Gradient design — i.e., the Control/Impact design was better able to pick out significant effects. The reduced resolving power of the Gradient design is probably related to the patchy distribution of local habitat types in most receiving environments, making it difficult to position sampling stations evenly along a gradient without locating some stations in slightly different habitat types. In addition, more exposed stations in the Gradient study design are being compared with less exposed stations at the far end of the gradient, rather than with entirely unexposed stations, as in the Control/Impact study design, so that Gradient designs would not be expected to be measuring the full magnitude of the effluent effects.

On the other hand, the Gradient design also has certain advantages. It can be better at detecting spatial patterns in effects, helpful in estimating the geographic extent of effects, and useful when rapid effluent dilution precludes extensive sampling within an exposure area with a relatively homogeneous effluent concentration.

A few key ways to improve Gradient study designs in future cycles include locating all sampling stations in comparable local habitat types to increase resolving power (this may require that stations are not spaced evenly along the gradient); extending the gradient far enough from the outfall so that the most distant sampling stations approximate the reference condition as much as possible; and expending greater sampling effort (greater number of sampling stations) on the two ends of the gradient, to increase the resolving power of the regression analyses and provide data appropriate for concurrent ANOVA comparisons, which are more directly comparable with results obtained from Control/Impact design studies.

8.4.2.2 Fish Population Survey

In the fish population survey, the use of small-bodied fish species in Cycle 2 proved to be successful, particularly at marine mills. Small-bodied fish species are easier to capture and are generally less mobile than larger fish species, so there is a greater likelihood of exposure to mill effluent. The most common freshwater finfish species used for Cycle 2 monitoring were the white sucker and longnose sucker, followed by the yellow perch.

Marine mills had much greater success in capturing finfish in Cycle 2 compared with Cycle 1. Mummichogs were used successfully for the purposes of the EEM survey, as they were caught in sufficient numbers and allowed for the measurement of all the EEM endpoints.

A greater number of mills made use of alternative methodologies during Cycle 2 to alleviate difficulties encountered with standard fish population surveys in marine/estuarine environments in Cycle 1. Mesocosms or caged bivalves were the alternatives of choice and provided useful data, showing that these methods can be used successfully for EEM studies.

8.4.3 Effect Size for Calculating Required Sampling Effort

The determination of appropriate predetermined effect sizes is an issue that has a number of important implications for the EEM program. Three kinds of effect sizes are described in this report — 1) a predetermined critical effect size, exceedence of which might require follow-up action; 2) the predetermined magnitude of effect that the EEM program would like to be able to detect and that is used to calculate required sampling effort; and 3) the magnitude of effect that is actually measured in a given study. The following discussion refers to the second kind of effect size.

8.4.3.1 Benthic Invertebrate Community Survey

One of the proposals under consideration for the benthic invertebrate component of the EEM program has been to set the predetermined effect size in terms of the degree of departure from the normal range of variability observed in reference areas. One measure of the normal range of variability that has been used in the literature is ±2 standard deviations about the mean for the reference community. An examination of the benthic invertebrate data from Cycle 2 shows that using two standard deviations as a predetermined effect size for calculating required sampling effort has some value for determining which mill exposure areas are showing the most pronounced effects.

8.4.3.2 Fish Population Survey

For the fish population survey, the Cycle 1 recommendation of a ±25% effect size remains an appropriate target effect size for gonad and liver weight, as well as weight at age, based on the Cycle 2 results. For condition, the target effect size should be closer to 10%, as this endpoint is less variable and does not respond as dramatically as the other endpoints to effluent exposure.

8.4.4 Data Presentation

It is recommended that only data used by mills for analysis be included in their electronic data submission. For Cycle 3, the use of the new pulp and paper EEM software should improve the quality of electronic data submitted.

It is also recommended that the mills present summary tables in the interpretative reports, including such summary statistics as arithmetic means, medians, standard deviations, standard errors, and minimum and maximum values. The inclusion of scatterplots is also recommended for ANCOVA/regression analyses. Histograms (with standard error bars) should be used to present the results of ANOVA analyses.

8.4.5 Cycle 2 Results

8.4.5.1 Benthic Invertebrate Community Survey

Effects on benthic invertebrates appear to have improved since Cycle 1. Effects are still widely present, but do not appear to be pronounced at most mills.

The national average response pattern for benthic invertebrates was one of mild to moderate eutrophication. Within rivers, the national average response pattern was one of mild to moderate eutrophication, with pronounced eutrophication at a few mills. In marine receiving environments, the national average response pattern for benthos was one of toxicity/smothering. Effects on benthic invertebrates in lakes and estuaries were intermediate between those in rivers and marine receiving environments.

8.4.5.2 Fish Population Survey

The dominant pattern found in Cycle 2 was a decrease in gonad weight and an increase in liver weight, condition factor, age, and weight at age for fish in the exposure area compared with reference fish. This pattern is consistent with a pattern of metabolic disruption (decrease in energy use) and nutrient enrichment (increase in energy storage). Some mills also had a pattern consistent with food limitation (decreases in the key endpoints) and chemical toxicity (increase in liver weight, decreases in condition and gonad weight).

8.4.5.3 Sublethal Toxicity Testing

Based on the data for sublethal toxicity tests summarized here, there has been a national improvement in effluent quality at Canadian pulp and paper mills between Cycle 1 and Cycle 2. Installation of secondary treatment between the two cycles was likely a major factor in lessening sublethal toxicity.

8.4.6 Ongoing Analyses and Interpretation

This national assessment synthesizes a number of the initial findings resulting from analyses of data provided by the EEM program. These analyses will continue, using both data already

collected as well as data provided by future cycles of the program. Where appropriate data are available, future issues to be addressed include:

- Changes in the magnitude and nature of effects from one cycle to the next. Quantitative analysis of temporal changes is key to evaluating the effectiveness of improvements in mill operations in mitigating negative effects on receiving environments.

- Compliance data. Given the way in which mill effluents are regulated, it is important to determine the relationship of compliance data to receiving environment effects, via comparisons with EEM field measurements.

- Effects of mill process and effluent treatment types. As more data become available, it may be possible to investigate the effects of some of the smaller mill subgroups that appear to be having quite pronounced environmental effects.

- Assessment of the relationship of EEM sublethal toxicity data to field data via a combination of meta-analyses and correlation analyses.

- Evaluation of regional patterns of effects.

- Influence of environmental and other co-variables on measured effects (e.g., effluent dilution, distance to outfall, variation in habitat types, water quality measures, co-occurring effluents, etc.).

- Benthic invertebrates:
  - Influence of taxonomic level of analysis on patterns of effects.
  - Change in magnitude and nature of effects from near to far field.
  - Identification of individual taxon changes in abundance having the greatest influence on changes in overall community structure in exposure areas.

- Fish:
  - Comparison of effects on large species with those on small species.
  - Similarities or differences in effects among species at different trophic levels.
  - Consistency in responses among studies for widely used sentinel species.

8.5 Overall Conclusions

The Canadian pulp and paper industry has invested billions of dollars towards research and technology aimed at reducing their emissions and improving the quality of their effluent. Today, most pulp and paper mills in Canada are equipped with both primary and secondary treatment systems, and some have tertiary treatment systems as well. In addition to advancements in pollution control, the industry continually works to modify its processes with the aim of achieving more effective pollution prevention. Environmental initiatives by the industry have resulted in the elimination of major pollutants like dioxins and furans in wastewater and a dramatic improvement in water quality. Together, these have led to dramatic improvements in the chemical composition and toxicity of the final effluent and have significantly reduced the

potential threat to the environment. The data from Cycle 2 were collected only a few years after many mills installed secondary treatment and already there appears to be evidence of reduced incidences of negative effects on benthic invertebrate communities.

While effluent quality has vastly improved since the PPER were promulgated; pulp and paper mills continue to have an impact on fish and benthos, underlining the importance of the ongoing collection of good data based on sound EEM studies. More monitoring is needed to assess the spatial extent and ecological importance of these observed effects, and Environment Canada will continue to work with industry and other stakeholders to better understand their significance.

References

Adams, S.M., W.D. Crumby, M.S. Greeley, Jr., L.R. Shugart, and C.F. Saylor. 1992. Responses of fish populations and communities to pulp mill effluents: A holistic assessment. Ecotoxicol. Environ. Saf. 24: 347–360.

Bailey, R.C., R.H. Norris, and T.B. Reynoldson. 2001. Taxonomic resolution of benthic macroinvertebrate communities in bioassessments. J. N. Am. Benthol. Soc. 20: 280–286.

Belbin, L. 1992. PATN pattern analysis package: technical reference. CSIRO, Lyneham, Australia.

Biermann, C.J. 1996. Handbook of pulping and papermaking. Academic Press, San Diego, California.

Bonsor, N., N. McCubbin, and J.B. Sprague. 1988. Kraft mill effluents in Ontario. Prepared for the Ontario Ministry of the Environment. 260 pp. (Ref. Type: Report 98).

Bothwell, M.L. 1992. Eutrophication of rivers by nutrients in treated kraft pulp mill effluent. Water Pollut. Res. J. Can. 27: 447–472.

Bothwell, M.L. and J.G. Stockner. 1979. Influence of secondarily treated kraft mill effluent on the accumulation rate of attached algae in experimental continuous flow-throughs. Can. J. Fish. Aquat. Sci. 37: 248–254.

Bowman, J.F. and R.C. Bailey. 1997. Does taxonomic resolution affect the multivariate description of the structure of freshwater benthic macroinvertebrate communities? Can. J. Fish. Aquat. Sci. 54: 1802–1807.

Chambers, P.A., S. Brown, J.M. Culp, R.B. Lowell, and A. Pietroniro. 2000a. Dissolved oxygen decline in ice-covered rivers on northern Alberta and its effects on aquatic biota. J. Aquat. Ecosyst. Stress Recov. 8: 27–38.

Chambers, P.A., A.R. Dale, G.J. Scrimgeour, and M.L. Bothwell. 2000b. Nutrient enrichment of northern rivers in response to pulp mill and municipal discharges. J. Aquat. Ecosyst. Stress Recov. 8: 53–66.

Chapman, G.A., D.L. Denton, and J.M. Lazorchak (eds.). 1995. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to west coast marine and estuarine organisms. Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C. 661 pp. (EPA 600/R-95/136).

Colby, P.J. 1984. Appraising the status of fisheries: rehabilitation techniques. *In* V.W. Cairns, P.V. Hodson, and J.O. Nriagu (eds.), Contaminant effects on fisheries. Adv. Environ. Sci. Technol. 16: 233–257.

Colby, P.J. and S.J. Nepszy. 1981. Variation among stocks of walleye (*Stizostedion vitreum vitreum*): management implications. Can. J. Fish. Aquat. Sci. 38: 1814–1831.

Courtenay, S.C., K. Munkittrick, H.M.C. Dupuis, R. Parker, and J. Boyd. 2002. Quantifying impacts of pulp mill effluent on fish in Canadian marine and estuarine environments: problems and progress. Water Qual. Res. J. Can. 37(1): 77–99.

Culp, J.M., R.B. Lowell, and K.J. Cash. 2000a. Integrating mesocosm experiments with field and laboratory studies to generate weight-of-evidence risk assessments for large rivers. Environ. Toxicol. Chem. 19: 1167–1173.

Culp, J.M., C.L. Podemski, and K.J. Cash. 2000b. Interactive effects of nutrients and contaminants from pulp mill effluents on riverine benthos. J. Aquat. Ecosyst. Stress Recov. 8: 67–75.

Dubé, M.G. and J.M. Culp. 1996. Growth responses of periphyton and chironomids exposed to biologically treated bleached kraft pulp mill effluent. Environ. Toxicol. Chem. 15: 2019–2027.

Dubé, M.G. and D.L. McLatchy. 2000. Endocrine responses of *Fundulus heteroclitus* to effluent from a bleached-kraft mill before and after installation of reverse osmosis treatment of a waste stream. Environ. Toxicol. Chem. 19: 2788–2796.

EC/DFO (Environment Canada/Department of Fisheries and Oceans). 1993. Interpretative guidance for the adult fish population survey requirement for environmental effects monitoring (EEM). Ottawa, Ontario.

Environment Canada. 1990. *Canadian Environmental Protection Act.* Priority Substances List Assessment Report No. 1. Polychlorinated dibenzodioxins and polychlorinated dibenzofurans (http://www.ec.gc.ca/dioxin/download/dioxin_assess_e.pdf).

Environment Canada. 1992a. Biological test method: test of reproduction and survival using the cladoceran *Ceriodaphnia dubia*. Environmental Protection Service, Environment Canada, Ottawa, Ontario (EPS 1/RM/21).

Environment Canada. 1992b. Biological test method: test of larval growth and survival using fathead minnows. Environmental Protection Service, Environment Canada, Ottawa, Ontario (EPS 1/RM/22).

Environment Canada. 1992c. Biological test method: growth inhibition test using the freshwater alga *Selenastrum capricornutum*. Environmental Protection Service, Environment Canada, Ottawa, Ontario (EPS 1/RM/25).

Environment Canada. 1992d. Biological test method: fertilization assay using echinoids (sea urchins and sand dollars). Environmental Protection Service, Environment Canada, Ottawa, Ontario (EPS 1/RM/27).

Environment Canada. 1997a. Toxicology expert working group report. National EEM Office, Science Policy and Environmental Quality Branch, Environment Canada, Ottawa, Ontario (EEM/1997/5).

Environment Canada. 1997b. Recommended procedure for adjusting salinity of effluent samples for marine sublethal toxicity testing conducted under environmental effects monitoring (EEM) programs. Environmental Technology Centre Report, Environmental Protection Service, Environment Canada, Ottawa, Ontario, July. 15 pp.

Environment Canada. 1997c. Fish population survey expert working group final report: Recommendations from Cycle 1 review. National EEM Office, Science Policy and Environmental Quality Branch, Environment Canada, Ottawa, Ontario (EEM/1997/6).

Environment Canada. 1997d. Benthic invertebrate community expert working group final report: Recommendations from Cycle 1 review. National EEM Office, Science Policy and Environmental Quality Branch, Environment Canada, Ottawa, Ontario (EEM/1997/7).

Environment Canada. 1998a. Biological test method: toxicity tests using early life stages of rainbow trout. 2nd edition. Environmental Protection Service, Environment Canada, Ottawa, Ontario (EPS 1/RM/28).

Environment Canada. 1998b. Pulp and paper technical guidance for aquatic environmental effects monitoring. National EEM Office, Science Policy and Environmental Quality Branch, Environment Canada, Ottawa, Ontario (EEM/1998).

Environment Canada. 1999a. Guidance document for implementing and interpreting single-species tests in environmental toxicology. Environmental Technology Centre, Environment Canada, Ottawa, Ontario (EPS 1/RM/34).

Environment Canada. 1999b. Canada's response to the recommendations in the Ninth Biennial Report of the International Joint Commission (http://www.on.ec.gc.ca/doc/ninth-ijc-response/dioxinfuran-e.html).

Environment Canada. 2002. Metal mining guidance document for aquatic environmental effects monitoring. National EEM Office, Science Policy and Environmental Quality Branch, Environment Canada, Ottawa, Ontario (http://www.ec.gc.ca/eem).

Faith, D.P., P.R. Minchin, and L. Belbin. 1987. Compositional dissimilarity as a robust measure of ecological distance. Vegetatio 69: 57–68.

Feder, H.M. and T.H. Pearson. 1988. The benthic ecology of Loch Linnhe and Loch Eil, a sea-loch system on the west coast of Scotland: V. Biology of the dominant soft-bottom epifauna and their interaction with the infauna. J. Exp. Mar. Biol. Ecol. 116: 99–134.

Felder, D.P., S.J. D'Surney, J.H. Rodgers, Jr., and T.L. Deardorff. 1998. A comprehensive environmental assessment of a receiving aquatic system near an unbleached kraft mill. Ecotoxicology 7: 313–324.

FPAC (Forest Products Association of Canada). 2001. FPAC web site: http://www.fpac.ca.

Gibbons, W.N. and K.R. Munkittrick. 1994. A sentinel monitoring framework for identifying fish population responses to industrial discharges. J. Aquat. Ecosyst. Health 3: 227–237.

Gibbons, W.N., K.R. Munkittrick, M.E. McMaster, and W.D. Taylor. 1998. Monitoring aquatic environments receiving industrial effluents using small fish species. 2. Comparisons between responses of trout-perch (*Percopsis omiscomaycus*) and white sucker (*Catostomus commersoni*) downstream of a pulp mill. Environ. Toxicol. Chem. 17(11): 2238–2245.

Glozier, N.E., J.M. Culp, T.B. Reynoldson, R.C. Bailey, R.B. Lowell, and L. Trudel. 2002. Assessing metal mine effects using benthic invertebrates for Canada's Environmental Effects Program. Water Qual. Res. J. Can. 37: 251–278.

Godfray, H.C.J. 2002. Challenges for taxonomy. Nature 417: 17–19.

Green, R.H. 1989. Power analysis and practical strategies for environmental monitoring. Environ. Res. 50: 195–205.

Gurevitch, J. and L.V. Hedges. 2001. Meta-analysis. Combining the results of independent experiments. *In* S.M. Scheiner and J. Gurevitch (eds.). Design and analysis of ecological experiments. Oxford University Press, New York, N.Y. pp. 347–369.

Hall, T.J., R.K. Halley, and L.E. Lafleur. 1991. Effects of biologically treated kraft mill effluent on cold water stream productivity in experimental stream channels. Environ. Toxicol. Chem. 10: 1051–1060.

Hansson, S. 1987. Effects of pulp and paper mill effluents on coastal fish communities in the Gulf of Bothnia, Baltic Sea. Ambio 16: 344–348.

Hedges, L.V. and I. Olkin. 1985. Statistical methods for meta-analysis. Academic Press, New York, N.Y.

Hodson, P.V., K.R. Munkittrick, R. Stevens, and A. Colodey. 1996. A tier-testing strategy for managing programs of environmental effects monitoring. Water Qual. Res. J. Can. 31: 215–224.

Kelso, J.R.M. 1977. Density, distribution, and movement of Nipigon Bay fishes in relation to a pulp and paper mill effluent. J. Fish. Res. Board Can. 34: 879–885.

Kilgour, B. 2001. Analysis of covariance of 2nd cycle adult fish population survey data. Environment Canada, Ottawa, Ontario.

Kovacs, T.G., J.S. Gifford, L.A. Tremblay, B.I. O'Conner, P.H. Martel, and R.H. Voss. 1995. The effects of a secondary-treated bleached kraft mill effluent on aquatic organisms as assessed by short-term and long-term laboratory tests. Ecotoxicol. Environ. Saf. 31: 7–22.

Lehtinen, K.-J. 1996. Biochemical responses in organisms exposed to effluents from pulp production: Are they related to bleaching?. *In* M.R. Servos, K.R. Munkittrick, J.H. Carey, and G.J. Van Der Kraak (eds.), Environmental fate and effects of pulp and paper mill effluents. St. Lucie Press, Delray, Florida. pp. 359–368.

Lenat, D.R. and V.H. Resh. 2001. Taxonomy and stream ecology — The benefits of genus- and species-level identifications. J. N. Am. Benthol. Soc. 20: 287–298.

Lowell, R.B. 1997. Discussion paper on critical effect size guidelines for EEM using benthic invertebrate communities. National EEM Office, Science Policy and Environmental Quality Branch, Environment Canada, Ottawa, Ontario (EEM/1997/8).

Lowell, R.B. 1999. Critical effect size guidelines as part of a weight-of-evidence approach for environmental effects monitoring. *In* S.L. Walker (ed.), Proceedings of a workshop on integrated approaches for interpreting environmental effects monitoring data. National EEM Office, Science Policy and Environmental Quality Branch, Environment Canada, Ottawa, Ontario. pp. 34–41 (EEM/1999/2).

Lowell, R.B. and J.M. Culp. 1999. Cumulative effects of multiple effluent and low dissolved oxygen stressors on mayflies at cold temperatures. Can. J. Fish. Aquat. Sci. 56: 1624–1630.

Lowell, R.B. and J.M. Culp. 2002. Implications of sampling frequency for detecting temporal patterns during environmental effects monitoring. Water Qual. Res. J. Can. 37: 119–132.

Lowell, R.B., J.M. Culp, and F.J. Wrona. 1995. Stimulation of increased short-term growth and development of mayflies by pulp mill effluent. Environ. Toxicol. Chem. 14: 1529–1541.

Lowell, R.B., J.M. Culp, F.J. Wrona, and M.L. Bothwell. 1996. Effects of pulp mill effluent on benthic freshwater invertebrates: food availability and stimulation of increased growth and development. *In* M.R. Servos, K.R. Munkittrick, J.H. Carey, and G.J. Van Der Kraak (eds.), Environmental fate and effects of pulp and paper mill effluents. St. Lucie, Delray Beach, Florida. pp. 525–532.

Lowell, R.B., J.M. Culp, and M.G. Dubé. 2000. A weight-of-evidence approach for northern river risk assessment: integrating the effects of multiple stressors. Environ. Toxicol. Chem. 19: 1182–1190.

Lowell, R.B., K. Hedley, and E. Porter. 2002. Data interpretation issues for Canada's Environmental Effects Monitoring Program. Water Qual. Res. J. Can. 37: 101–117.

Mayack, D.T. and J.S. Waterhouse. 1983. The effects of low concentrations of particulates from paper mill effluent on the macroinvertebrate community of a fast-flowing stream. Hydrobiologia 107: 271–282.

McMaster, M.E., G.J. Van Der Kraak, C.B. Portt, K.R. Munkittrick, P.K. Sibley, I.R. Smith, and D.G. Dixon. 1991. Changes in hepatic mixed-function oxygenase (MFO) activity, plasma steroid levels and age at maturity of a white sucker (*Catostomus commersoni*) population exposed to bleached kraft pulp mill effluent. Aquat. Toxicol. 21: 199–218.

Munkittrick, K.R. and D.G. Dixon. 1989a. A holistic approach to ecosystem health and assessment using fish population characteristics. Hydrobiologia 188/189: 122–135.

Munkittrick, K.R. and D.G. Dixon. 1989b. Use of white sucker (*Catostomus commersoni*) populations to assess the health of aquatic ecosystems exposed to low-level contaminant stress. Can. J. Fish. Aquat. Sci. 46: 1455–1462.

Munkittrick, K.R. and M.E. McMaster. 2000. Effects-driven assessment of multiple stressors using fish populations. *In* S.A. Ferene and J.A. Foran (eds.), Multiple stressors in ecological risk and impact assessment: Approaches to risk estimation. Society of Environmental Toxicology and Chemistry (SETAC) Press, Pensacola, Florida. pp. 27–65.

Munkittrick, K.R., C.B. Portt, G.J. Van Der Kraak, I.R. Smith, and D.A. Rokosh. 1991. Impacts of bleached kraft mill effluent on population characteristics, liver MFO activity, and serum steroid levels of a Lake Superior white sucker (*Catostomus commersoni*) population. Can. J. Fish. Aquat. Sci. 48: 1371–1380.

Munkittrick, K.R., G.J. Van Der Kraak, M.E. McMaster, C.B. Portt, M.R. van den Heuvel, and M.R. Servos. 1994. Survey of receiving water environmental impacts associated with discharges from pulp mills. 2. Gonad size, liver size, hepatic EROD activity and plasma sex steroid levels in white sucker. Environ. Toxicol. Chem. 13: 1089–1101.

Munkittrick, K.R., M.R. Servos, J.H. Carey, and G.J. Van Der Kraak. 1997. Environmental impacts of pulp and paper wastewater: Evidence for a reduction in environmental effects at North American pulp mills since 1992. Water Sci. Technol. 35: 329–338.

Munkittrick, K.R., M.E. McMaster, L.H. McCarthy, M.R. Servos, and G.J. Van Der Kraak. 1998. An overview of recent studies on the potential of pulp mill effluents to impact reproductive function in fish. J. Toxicol. Environ. Health, Part B 1: 101–125.

Munkittrick, K.R., M.E. McMaster, G. Van Der Kraak, C. Portt, W.N. Gibbons, A. Farwell, and M. Gray. 2000. Development of methods for effects-based cumulative effects assessment using fish populations: Moose River Project. Society of Environmental Toxicology and Chemistry (SETAC) Press, Pensacola, Florida.

Munkittrick, K.R., S.A. McGeachy, M.E. McMaster, and S.C. Courtenay. 2002. Review of cycle 2 freshwater fish studies from the pulp and paper environmental effects monitoring program. Water Qual. Res. J. Can. 37(1): 49–77.

OECD (Organisation for Economic Co-operation and Development). 1999. Canada. *In* Environmental requirements for industrial permitting: Country profiles on the pulp and paper sector: Part 2. Working Party on Pollution Prevention and Control, Environmental Directorate, OECD. pp. 41–77.

Owens, J.W. 1991. The hazard assessment of pulp and paper effluents in the aquatic environment: A review. Environ. Toxicol. Chem. 10: 1511–1540.

Pearson, T.H. and R. Rosenberg. 1978. Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. Oceanogr. Mar. Biol. Annu. Rev. 16: 229–311.

Rosenberg, M.S., D.C. Adams, and J. Gurevitch. 2000. MetaWin: Statistical software for meta-analysis. Version 2.0. Sinauer Associates, Sunderland, Massachusetts.

Scrimgeour, G.J. 1989. Effects of bleached kraft mill effluent on macroinvertebrate and fish populations in weedbeds in a New Zealand hydro-electric lake. N. Z. J. Mar. Freshwater Res. 23: 373–379.

Scroggins, R.P., J.A. Miller, A.I. Borgmann, and J.B. Sprague. 2002. Sublethal toxicity findings by the pulp and paper industry for cycles 1 and 2 of the environmental effects monitoring program. Water Qual. Res. J. Can. 37(1): 21–48.

Servos, M.R., S. Huestis, D.M. Whittle, G.J. Van Der Kraak, and K.R. Munkittrick. 1994. Survey of receiving water environmental impacts associated with discharges from pulp mills. 3. Polychlorinated dioxins and furans in muscle and liver of white sucker (*Catostomus commersoni*). Environ. Toxicol. Chem. 13: 1103–1115.

Servos, M.R., K.R. Munkittrick, J.H. Carey, and G.J. Van Der Kraak (eds.). 1996. Environmental fate and effects of pulp and paper mill effluents. St. Lucie Press, Delray, Florida.

Sibley, P.K., D.G. Dixon, and D.R. Barton. 1998. Environmental assessment of benthic impacts associated with pulp mill discharges. II. Distribution of sediment EOX in relation to environmental factors. Arch. Environ. Contam. Toxicol. 34: 158–166.

Sibley, P.K., D.G. Dixon, and D.R. Barton. 2001. Impact of bleached kraft pulp mill effluent on the nearshore benthic community of Jackfish Bay, Lake Superior. Water Qual. Res. J. Can. 36: 815–833.

Smith, B. and J.B. Wilson. 1996. A consumer's guide to evenness indices. Oikos 76: 70–82.

Thomas, G.P. and N. Munteanu. 1997. Benthic community populations near two adjacent northern pulpmill discharges. Water Sci. Technol. 35: 381–388.

USEPA (U.S. Environmental Protection Agency). 1994. Short-term methods for estimating the chronic toxicity of effluent and receiving waters to marine and estuarine organisms. 2nd edition. USEPA, Cincinnati, Ohio (EPA/600/4-91/003).

USEPA (U.S. Environmental Protection Agency). 1998. Pulp and paper NESHAP: A plain English description. USEPA, Washington, D.C. (EPA-456/R-98-008).

VanderWal, J. 1977. Relation between Nipigon Bay benthic macroinvertebrates and selected aspects of their habitat. J. Fish. Res. Board Can. 34: 824–829.

National Water Research Institute
Environment Canada

Canada Centre for Inland Waters
P.O. Box 5050
867 Lakeshore Road
Burlington, Ontario
L7R 4A6 Canada

National Hydrology Research Centre
11 Innovation Boulevard
Saskatoon, Saskatchewan
S7N 3H5 Canada

Water Quality Monitoring Branch
351 St. Joseph Boulevard
Gatineau, Quebec
K1A 0H3

NATIONAL WATER RESEARCH INSTITUTE
INSTITUT NATIONAL DE RECHERCHE SUR LES EAUX

Institut national de recherche sur les eaux
Environnement Canada

Centre canadien des eaux intérieures
Case postale 5050
867, chemin Lakeshore
Burlington, Ontario
L7R 4A6 Canada

Centre national de recherche en hydrologie
11, boul. Innovation
Saskatoon, Saskatchewan
S7N 3H5 Canada

Direction de la surveillance de la qualité de l'eau
351 boul. St-Joseph
Gatineau, Québec
K1A 0H3

Environment Canada Environnement Canada

Canada